现代数学基础丛书 145

现代统计模型

薛留根 著

科学出版社

(O-4813.0101)

www.sciencep.com



科学出版中心 数理分社
电话：(010) 64033664
Email：math-phy@mail.sciencep.com
网 址：http://www.math-phy.cn

销售分类建议：高等数学

定 价：88.00 元

现代数学基础丛书 145

现代统计模型

薛留根 著

科学出版社
北京

内 容 简 介

本书阐述了半参数回归模型的统计理论和方法，所考虑的模型包括部分线性模型、单指标模型、变系数模型和可加模型等. 这些模型对复杂数据分析起着重要作用. 本书在取材上侧重内容的科学性和应用性，体现学术思想；在写作上注重阐述方法论、模拟计算和实例分析；在结构上安排每个模型为一章. 本书的内容不仅为从事该领域的科研人员提供了尽可能全的资料，又为实际应用者提供了一些数据分析的方法，同时也为想全面了解现代统计模型的读者提供参考读物.

本书可以作为高等院校有关专业的教学用书，同时也对高等院校和科研机构的研究人员、工程技术人员和研究生有参考价值.

图书在版编目(CIP)数据

现代统计模型/薛留根著. —北京：科学出版社，2012
ISBN 978-7-03-035020-6
(现代数学基础丛书；145)
Ⅰ. ①现… Ⅱ. ①薛… Ⅲ. ①统计模型-研究 Ⅳ. ①C8
中国版本图书馆 CIP 数据核字（2012）第 133253 号

责任编辑：陈玉琢／责任校对：朱光兰
责任印制：钱玉芬／封面设计：王 浩

科学出版社出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com
骏杰印刷厂印刷
科学出版社发行 各地新华书店经销
*
2012 年 6 月第 一 版 开本：B5(720×1000)
2012 年 6 月第一次印刷 印张：26 1/4
字数：500 000
定价：88.00 元
(如有印装质量问题，我社负责调换)

《现代数学基础丛书》序

对于数学研究与培养青年数学人才而言, 书籍与期刊起着特殊重要的作用. 许多成就卓越的数学家在青年时代都曾钻研或参考过一些优秀书籍, 从中汲取营养, 获得教益.

20世纪70年代后期, 我国的数学研究与数学书刊的出版由于文化大革命的浩劫已经破坏与中断了 10 余年, 而在这期间国际上数学研究却在迅猛地发展着. 1978 年以后, 我国青年学子重新获得了学习、钻研与深造的机会. 当时他们的参考书籍大多还是50年代甚至更早期的著述. 据此, 科学出版社陆续推出了多套数学丛书, 其中《纯粹数学与应用数学专著》丛书与《现代数学基础丛书》更为突出, 前者出版约40卷, 后者则逾80卷. 它们质量甚高, 影响颇大, 对我国数学研究、交流与人才培养发挥了显著效用.

《现代数学基础丛书》的宗旨是面向大学数学专业的高年级学生、研究生以及青年学者, 针对一些重要的数学领域与研究方向, 作较系统的介绍. 既注意该领域的基础知识, 又反映其新发展, 力求深入浅出, 简明扼要, 注重创新.

近年来, 数学在各门科学、高新技术、经济、管理等方面取得了更加广泛与深入的应用, 还形成了一些交叉学科. 我们希望这套丛书的内容由基础数学拓展到应用数学、计算数学以及数学交叉学科的各个领域.

这套丛书得到了许多数学家长期的大力支持, 编辑人员也为其付出了艰辛的劳动. 它获得了广大读者的喜爱. 我们诚挚地希望大家更加关心与支持它的发展, 使它越办越好, 为我国数学研究与教育水平的进一步提高做出贡献.

杨　乐

2003 年 8 月

前　言

在实际问题中, 人们常常遇到各种各样的数据, 如独立数据、纵向数据、缺失数据、删失数据、测量误差数据等. 统计模型是对数据内在规律性的数学描述. 由于所研究的客观现象越来越复杂, 要求也越来越精细, 回归模型已从线性模型发展到各种非线性回归模型、非参数和半参数回归模型. 随着科学技术的快速发展, 半参数回归模型对复杂数据的分析起着重要应用, 其中包括部分线性模型、单指标模型、变系数模型和可加模型等. 这类模型通过引入参数分量和非参数函数, 能够真正组合数据的相关信息, 从而提高了模型的拟合优度. 半参数回归模型是统计学中重要的模型, 它们在信息科学、生物医学、金融工程、经济管理、质量控制、交通工程、能源与环境、人文与社会科学等领域都有着相当广泛的应用. 近年来, 国内外许多学者致力于对这些模型的研究, 现已成为统计学的热门研究课题之一.

本书系统阐述了半参数回归模型的统计推断方法, 主要内容包括估计方法、模型检验和变量选择等. 本书不但介绍了各种模型的发展动态, 而且展示了最新的研究成果. 本书的内容不仅为从事该领域的科研人员提供了尽可能全的资料, 又为实际应用者提供了一些数据分析方法, 同时也为想全面了解现代统计模型的读者提供参考读物.

本书根据作者多年从事科研工作所积累的大量资料、并结合本人的研究成果撰写而成. 在取材上精心提炼, 既考虑到了深度与广度, 又顾及了内容的科学性和应用性, 体现学术思想. 在写作上注重阐述方法论、模拟计算和实例分析, 省略或简化理论证明. 本书结构是把每个模型作为一章, 读者可以独立阅读每一章内容. 初学者可以避开阅读理论部分.

本书得到了国家自然科学基金 (11171012)、北京市优秀博士学位论文指导教师科技项目 (20111000503) 和北京市教育委员会科技计划项目 (KM201110005029) 的资助, 作者谨在此表示感谢.

薛留根

2012 年 3 月

符 号 表

iid	独立同分布
$\mathbf{R}^p$	p 维 Euclid 空间
$\|\|\cdot\|\|$	Euclid 模
A^{T}	向量或矩阵 A 的转置
$\mathrm{tr}(A)$	矩阵 A 的迹
$I_B(\cdot)$或$I(B)$	集合 B 的示性函数
$A\otimes B$	矩阵 A 与 B 的 Kronecker 乘积
$A^{\otimes 2}$	AA^{T}
$\mathrm{diag}\{a_1,\cdots,a_n\}$	由元素 $a_1,\cdots,a_n$ 组成的对角矩阵
$\xrightarrow{D}$	依分布收敛
$\xrightarrow{P}$	依概率收敛
a.s.	几乎处处收敛
$\equiv$	“定义为”或“记为”
$y=O(1)$	y 是有界变量
$y=o(1)$	y 是无穷小量
$\xi_n=o_p(\eta_n)$	对任一 $\varepsilon>0$, 有 $P\{\|\|\xi_n\|\|\geqslant\varepsilon\|\|\eta_n\|\|\}\to 0$
$\xi_n=o_P(1)$	ξ_n 依概率收敛到 0
$\xi_n=O_P(1)$	ξ_n 依概率有界
$f'(t)$	函数 $f(t)$ 的一阶导数
$N(\mu,\Sigma)$	均值为 μ, 协方差阵为 Σ 的正态分布
χ_p^2	自由度为 p 的 χ^2 分布
$\chi_p^2(1-\alpha)$	自由度为 p 的 χ^2 分布的 $1-\alpha$ 分位数
$z_{1-\alpha/2}$	标准正态分布的 $1-\alpha/2$ 分位数
x_+	变量 X 的正部
c	不依赖于 n 的正的常数, 在不同地方可以表示不同的值

目　录

第1章　预备知识

本章介绍一些预备知识, 以便读者阅读后面各个章节. 本章的主要内容包括: 回归模型简介、光滑方法、自动选择光滑参数、经验似然、惩罚方法、数据集和符号表等.

1.1　回归模型简介

回归模型是统计学中发展较早、理论丰富并且应用性强的统计模型. 自 Galton 于 1886 年首次提出回归模型以来, 这类模型一直倍受人们关注. 在过去的一个多世纪里, 回归模型被广泛地应用于工业、农业、气象、地质、经济管理以及医药卫生等各个领域, 取得了丰富的理论和应用成果. 同时, 由于实际应用的需要, 为了更加接近现实, 更好地解释数据, 回归模型也一直处于不断发展进步之中, 其模型由初期的参数回归模型发展到非参数回归模型, 20 世纪 80 年代以来又兴起了半参数回归模型, 其理论研究和应用领域都在不断深入和扩大.

1.1.1　参数回归模型

参数回归模型具有形式

$$Y_i = f(X_i;\beta) + \varepsilon_i, \quad i = 1, 2, \cdots, n,$$

其中 f 为已知函数, β 为未知的有限维参数, ε_i 为均值为 0 的随机误差. 这类模型包括大家熟知的线性模型、广义线性模型和非线性回归模型. 参数回归模型有一个共同的特点: 函数形式假定是已知的, 未知的仅仅是其中有限个参数. 由于参数回归模型容易处理, 且对其研究已有相当长的历史, 因而已形成一套成熟的理论和方法.

参数回归模型通常由经验和历史资料对回归函数提供了大量的额外信息, 因而当模型假定成立时, 其推断有较高的精度. 例如熟知的线性模型的最小二乘估计, 其方差有 $O(n^{-1})$ 的阶. 但响应变量与协变量之间的函数关系形式不总是已知的, 在许多实际应用中, 没有证据表明已知函数关系的存在. 因而, 参数回归模型在实际应用中往往存在模型的设定误差, 当模型假定不成立时, 基于假定模型所作的统计推断其表现可能很差, 甚至导致错误的结论.

为了减少参数回归的模型偏差, 统计学家提出了一个假设更宽松自由的模型 —— 非参数回归模型.

1.1.2　非参数回归模型

设响应变量 Y 是随机变量, 协变量 X 是随机变量或非随机变量. 给定随机样本 $\{(X_i, Y_i), 1 \leqslant i \leqslant n\}$, 可建立如下回归模型:

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n, \tag{1.1.1}$$

其中 $m(\cdot)$ 是未知的回归函数, ε_i 为随机误差. 假定 $\varepsilon_1, \cdots, \varepsilon_n$ 相互独立同分布, 并满足:

(1) X 为非随机时: $E(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2 < \infty$, 此时 $E(Y_i) = m(X_i)$.

(2) X 为随机时: $E(\varepsilon_i|X_i) = 0$, $\mathrm{Var}(\varepsilon_i|X_i) = \sigma^2(X_i) < \infty$, 此时 $E(Y_i|X_i) = m(X_i)$.

非参数回归模型的特点是: 回归函数的形式可以任意, 对协变量和响应变量的分布也很少限制, 因而有较大的适应性和稳健性. 非参数回归模型是基于数据的模型, 它假定响应变量和协变量之间的关系形式未知, 由观测数据本身对整个回归函数进行估计, 是较参数回归模型更符合现实的模型.

回归函数解释了 X 与 Y 之间的相互关系. 回归分析的目的是对于未知函数 $m(x)$, 给出一个合理的近似值, 称之为 $m(x)$ 的估计. 构造 $m(x)$ 的估计量的方法称为非参数回归方法, 又称为光滑方法, 与参数回归方法相反, 不采用现成的数学函数作为模型, 在统计领域中是较新的方法. 用非参数回归估计回归曲线具有以下特点: 第一, 关于两个变量关系的探索是开放式的, 不套用任何现成的数学函数. 第二, 所拟合的曲线可以很好地描述变量之间关系的细微变化. 第三, 非参数回归提供的是万能的拟合曲线, 不管多么复杂的曲线关系都能进行成功拟合. 因此, 非参数估计方法在分析未知回归函数关系时提供了一个灵活的工具.

与传统回归拟合相比, 非参数回归则显得灵活机动, 可以对同一数据进行多次拟合, 以探索数据中可能隐藏的某种关系. 这是普通的回归拟合和散点图绝对做不到的. 而且, 非参数拟合可以对数据中的任何模式或变量间任何一种曲线关系进行拟合, 而传统回归却只能对个别的数据模式 (直线、二次曲线等) 进行拟合. 实际上, 非参数回归拟合往往会带来意想不到的结果, 会极富戏剧性地改变人们对数据进行进一步分析的方向, 得到更深刻的结论.

在非参数回归模型中, 回归函数既可以是一元的, 也可以是多元的. 理论上讲, 一元非参数回归的估计方法可以直接推广到多元非参数回归, 但是有一个很棘手的问题, 即维数问题. 由于非参数回归估计方法本质上讲都是局部估计或局部光滑, 要想使 $m(x)$ 在 x 点得到比较精确的估计, 就必须使得 x 的邻域包含足够多的数据. 当 x 的维数增大时, 局部光滑所需要的数据点个数成指数倍增加. 例如, 如果一个局部邻域沿着每一个坐标轴包含 10 个数据点, 那么在相应的 p 维邻域就需要

10^p 个数据点. 而实际上, 高维数据具有内部稀疏性, 即随着维数 p 的增加, 一个局部邻域所包含的数据点个数在整个样本中所占的比例越来越小. 为了获取非参数光滑所必须的足够多的数据, 可以有两种选择, 或者增大带宽, 或者增加样本数目. 增大带宽必然导致估计的偏差增加, 而要获取数目非常多的样本在许多情况下是不实际的. 由此可见, 当协变量 X 的维数增加时, 多元非参数回归估计的精度下降很快, 人们称这种现象为"维数灾祸"(curse of dimensionality).

为了克服"维数灾祸"问题, 统计学家在寻找既能达到数据降维同时又能保留非参数光滑优点的方法方面已经作了许多研究, 提出了多种降维模型. 主要的降维模型为半参数回归模型.

1.1.3 半参数回归模型

本书仅介绍几种重要的半参数回归模型, 其中包括: 部分线性模型、单指标模型、部分线性单指标模型、变系数模型、部分线性变系数模型、单指标变系数模型、部分变系数单指标模型和可加模型等. 在以后各章节中再对这些模型作一一阐述.

对于半参数回归模型的研究, 主要集中在统计推断方法和大样本性质上. 研究的基本问题大致可分为五个方面.

(1) 估计量的构造方法和统计性质. 包括回归参数、模型误差方差和非参数回归函数的估计方法的研究, 并研究估计量的相合性、渐近正态性及其收敛速度等. 对参数分量的估计方法有: 最小二乘法、广义估计方程法、广义矩方法、M 估计法等. 对非参数分量的估计方法包括: 核光滑、局部多项式光滑、样条光滑、小波光滑等.

(2) 回归参数的置信域以及非参数回归函数置信带的构造方法. 构造置信域的常用的方法有经验似然和 bootstrap 法等. 这类问题的研究给出了估计的精度, 具有重要性.

(3) 模型检验和拟合优度检验. 主要研究各种检验统计量的构造方法及其统计性质, 如广义似然比检验、Wald 检验、经验似然比检验等. 这类问题的研究为实际应用提供了理论依据.

(4) 变量选择或模型选择. 常用的变量选择方法包括: LASSO(least absolute shrinkage and selection operator) 和 SCAD(smoothly clipped absolute deviation) 及其各种变形. 这类问题是目前统计学研究的热点课题之一.

(5) 估计的稳健性. 所谓稳健性, 直观意义是说, 当假设模型与真实模型相符时估计量具有良好的性质; 当假设模型与真实模型稍许偏离时, 其性能所受影响较小; 当假设模型与真实模型发生严重偏离时, 其性能也还过得去. 如参数的 M 估计就是一种稳健估计.

1.2 光滑方法

本节概述回归光滑方法, 主要包括: 核光滑、局部多项式光滑、样条光滑、权函数法等. 此处介绍的光滑方法着眼于单变量回归估计. 这主要是出于如下两种原因: 第一, 对于高维的变量 X 表达式将变得极为复杂; 第二, 在以后各章讨论各种回归模型时, 主要使用单变量光滑方法.

1.2.1 核光滑

Nadaraya 和 Watson 于 1964 年分别提出了非参数回归函数 $m(x)$ 的核估计, 文献中也称之为核光滑 (kernel smoothing). 其思路如下: 选定 $\mathbf{R}$ 上的函数 $K(\cdot)$ 和正的常数列 $h = h_n$, 记 $K_h(\cdot) = h^{-1}K(\cdot/h)$, 定义

$$\hat{m}_{\rm NW}(x) = \sum_{i=1}^{n} W_{ni}^{K}(x)Y_i, \tag{1.2.1}$$

其中

$$W_{ni}^{K}(x) = K_h(X_i - x)\Big/\sum_{j=1}^{n} K_h(X_j - x), \tag{1.2.2}$$

则称 $\hat{m}_{\rm NW}(x)$ 为 $m(x)$ 的 Nadaraya-Watson 估计, 简称为 N-W 估计或核估计; 称 $K(\cdot)$ 为核函数, h 为带宽. 通常取核函数 $K(\cdot)$ 为概率密度函数. 常用的核函数有:

(1) 均匀核: $K(u) = \dfrac{1}{2}I(|u| \leqslant 1)$,

(2) Epanechnikov 核 (或抛物线核): $K(u) = \dfrac{3}{4}(1 - u^2)I(|u| \leqslant 1)$,

(3) 四次核: $K(u) = \dfrac{15}{16}(1 - u^2)^2 I(|u| \leqslant 1)$,

(4) 高斯核: $K(u) = \dfrac{1}{\sqrt{2\pi}}{\rm e}^{-\frac{u^2}{2}}$.

通常称带宽 h 为光滑参数. 例如, $K(\cdot)$ 为均匀核时, $\hat{m}_{\rm NW}(x)$ 是集中在 x 附近一个邻域的样本的加权平均, 而 h 正好是这个邻域的宽度. 当 h 大时, 参与平均的样本就多, 会提高估计的精度, 但可能会增大偏差. 当 h 小时则正好相反. 因此, h 的选取至关重要. 关于它的选取问题将在 1.3 节解决.

在一些正则条件下, N-W 估计是相合的, 即当 $h \to 0$ 且 $nh \to \infty$ 时,

$$\hat{m}_{\rm NW}(x) \xrightarrow{P} m(x).$$

相对于正确指定下的参数模型而言, 非参数估计是有偏的. 偏差 ${\rm bias}(\hat{m}_{\rm NW}) = E[\hat{m}_{\rm NW}(x) - m(x)]$ 可以通过减小带宽而减小, 但 $\hat{m}_{\rm NW}(x)$ 的方差将增大. 非参数回归的技巧是如何平衡偏差和方差的值使得所需要的结果达到基本合理.

经过推导可以得到 N-W 估计的渐近偏差为

$$\mathrm{bias}(\hat{m}_{\mathrm{NW}}(x)) = \frac{1}{2}\kappa_2 h^2 \left[\frac{2m'(x)f'(x)}{f(x)} + m''(x)\right] + O(h^2), \tag{1.2.3}$$

其中 $\kappa_2 = \int u^2 K(u)\mathrm{d}u$, $f(x)$ 为 X 的概率密度函数. 渐近方差为

$$\mathrm{var}(\hat{m}_{\mathrm{NW}}(x)) = \frac{1}{nh}\frac{\sigma^2(x)}{f(x)}\int K^2(u)\mathrm{d}u + O(n^{-1}h^{-1}). \tag{1.2.4}$$

关于估计量 $\hat{m}_{\mathrm{NW}}(x)$ 的更精细的研究可参考文献 (Härdle, 1990) 或 (Pagan and Ullah, 1999). N-W 估计 (1.2.2) 可直接推广到多维的情况, 其中多维核可取为乘积核. 渐近偏差和渐近方差的类似的结果亦成立. 参看文献 (Härdle et al., 2004).

当对 (1.2.1) 式给出的估计量求导数和获得它的渐近性质时, 它的随机分母是不方便的. 假设数据已经按照 X 变量排序. Gasser 和 Müller(1979) 提出了下列估计量:

$$\hat{m}_{\mathrm{GM}}(x) = \sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K_h(u-x)\mathrm{d}uY_i, \tag{1.2.5}$$

其中 $s_i = (X_i + X_{i+1})/2$, $X_0 = -\infty$, $X_{n+1} = +\infty$. 称 $\hat{m}_{\mathrm{GM}}(x)$ 为 G-M 估计. 注意到 (1.2.5) 式中权的和是 1 且没有分母. 该估计量原先是对等间隔设计提出的, 但也能用于非等间隔设计. Müller(1988,1991) 给出了估计量 $\hat{m}_{\mathrm{GM}}(x)$ 的详细讨论.

从函数逼近的观点看, N-W 估计和 G-M 估计都可用局部常数逼近. 假设 $m(\cdot)$ 用一个常数 a 局部逼近, 那么利用局部最小二乘回归可得到估计量

$$\hat{a} = \arg\min_a \sum_{i=1}^{n}(Y_i - a)^2 w_i = \sum_{i=1}^{n} w_i Y_i \Big/ \sum_{i=1}^{n} w_i. \tag{1.2.6}$$

由此可以看出, N-W 估计和 G-M 估计的权的形式分别是 $w_i = K_h(X_i - x)$ 和 $w_i = \int_{s_{i-1}}^{s_i} K_h(u-x)\mathrm{d}u$.

1.2.2 局部多项式光滑

由于核估计存在边界效应, 即核估计在边界处收敛于实际函数的速度慢于在内点处的收敛速度, 而且核估计是局部加权平均, 其偏差较大, 它的偏差还与协变量的密度函数有关. 下面提出局部多项式估计.

设回归函数 $m(\cdot)$ 在 x 的邻域内有连续的 p 阶导数, 那么由 Taylor 展式可得下列近似公式:

$$m(z) \approx \sum_{j=0}^{p}\frac{m^{(j)}(x)}{j!}(z-x)^j \equiv \sum_{j=0}^{p}\beta_j(z-x)^j, \tag{1.2.7}$$

其中 z 为 x 邻域内的点. 从统计建模的观点上看, (1.2.7) 式是用一个简单的多项式模型局部的建模. 因此, 求 $m(x)$ 的估计量可归结为二次极小化问题

$$\min_{\beta\in\mathbf{R}^{p+1}}\sum_{i=1}^{n}\left\{Y_i-\sum_{j=0}^{p}\beta_j(X_i-x)^j\right\}^2K_h\left(X_i-x\right), \tag{1.2.8}$$

其中 $\boldsymbol{\beta}=(\beta_0,\beta_1,\cdots,\beta_p)^{\mathrm{T}}$ 为系数向量, $K_h(\cdot)=h^{-1}K(\cdot/h)$, $K(\cdot)$ 为核函数, h 为带宽. 显然, 这个极小化问题的结果是一个具有权 $K_h(X_i-x)$ 的加权最小二乘估计. 用 $\hat{\beta}=(\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_p)^{\mathrm{T}}$ 表示极小化问题 (1.2.8) 的解. 直接计算可得

$$\hat{\beta}=(\boldsymbol{X}^{\mathrm{T}}\mathbf{W}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{Y}, \tag{1.2.9}$$

其中 $\boldsymbol{Y}=(Y_1,\cdots,Y_n)^{\mathrm{T}}$, $\boldsymbol{W}=\mathrm{diag}\{K_h(X_1-x),\cdots,K_h(X_n-x)\}$ 为 $n\times n$ 对角矩阵,

$$\boldsymbol{X}=\begin{pmatrix}1 & X_1-x & \cdots & (X_1-x)^p\\ \vdots & \vdots & & \vdots\\ 1 & X_n-x & \cdots & (X_n-x)^p\end{pmatrix},$$

用 $e_{\nu+1}$ 表示 $p+1$ 维单位向量, 其第 $\nu+1$ 个位置的元素为 1, 其他元素全为 0. 那么 $\hat{\beta}_\nu=e_{\nu+1}^{\mathrm{T}}\hat{\beta}$, $\nu=0,1,\cdots,p$. 因此, $m^{(\nu)}(x)$ 的估计量为 $\hat{m}_\nu(x)=\nu!\hat{\beta}_\nu$. 当 x 在适当的估计范围内变化时, 通过实施上述局部多项式回归可以得到整个曲线 $\hat{m}_\nu(x)$.

在实际中要求选择 p. 从渐近性质可知奇数阶多项式胜过偶数阶多项式. 一个通常使用的情况是局部线性光滑, 即 $p=1$. 称估计量 $\hat{m}_0(x)$ 为局部线性估计, 并简记为 $\hat{m}_{\mathrm{LL}}(x)$. 经过计算可以得到

$$\hat{m}_{\mathrm{LL}}(x)=\sum_{i=1}^{n}W_{ni}^{\mathrm{L}}(x)Y_i,$$

其中

$$W_{ni}^{\mathrm{L}}(x)=n^{-1}\frac{K_h(X_i-x)[S_{n,2}(x)-(X_i-x)S_{n,1}(x)]}{S_{n,0}(x)S_{n,2}(x)-S_{n,1}^2(x)}, \tag{1.2.10}$$

$$S_{n,j}(x)=\frac{1}{n}\sum_{i=1}^{n}(X_i-x)^jK_h(X_i-x),\quad j=0,1,2.$$

对局部线性估计量 $\hat{m}_{\mathrm{LL}}(x)$, 渐近方差与 (1.2.4) 式给出的核估计量的渐近方差一样. 渐近偏差具有形式

$$\mathrm{bias}(\hat{m}_{\mathrm{LL}}(x))=\frac{1}{2}\kappa_2h^2m''(x)+O(h^2). \tag{1.2.11}$$

(1.2.11) 式与 (1.2.3) 式比较可以发现一个异常的差别：偏差不依赖于密度，即它是自适应设计，见文献 (Fan, 1992). 此外，当 $m(\cdot)$ 是线性函数时，偏差消失. 因此，当设计点稀疏时局部线性估计优于核估计. 局部线性估计的另一个优越性是它的偏差和方差在 $m(x)$ 的边界和内部是同阶的，因此不需要在边界点处用特殊的权函数来减少边界效应. 在实际中，这可以在设计的边界点处改良估计量的性质.

局部多项式估计是线性估计类中的最佳估计，它有几个吸引人的特点. 例如，它有好的最小最大性质；它可适用于各种设计，如随机设计和固定设计等；它容易解释、实施并适应于导数的估计等. 核估计是在 x 的一个局部邻域内拟合一个常数的特殊情况. 与核估计相比较，局部多项式估计的一个显著特点是它不存在边界效应问题，即在边界点处也有很好的性能. 此外，局部多项式估计在估计出回归函数的同时也估计出了回归函数的各阶导函数. 这些相关的工作可以查找文献 (Fan, 1993; Fan and Gijbels, 1995, 1996; Fan et al., 1995; Ruppert, 1997).

1.2.3 样条光滑

所谓样条就是拟合分段多项式，用样条逼近函数曲线的方法称为样条方法. 逼近函数的导数的间断点称为 "节". 样条实际上可用来逼近任何光滑函数，至少当节点数目充分大时可以做到这一点. 这个特性使得样条能很好地适合非参数回归问题. 关于样条逼近及其统计应用的参考文献有 (Eubank, 1988; Wahba, 1990; Green and Silverman, 1994) 等. 本节针对模型 (1.2.1) 介绍两种样条方法：多项式样条和光滑样条.

1. 多项式样条

假设我们打算用一个样条函数逼近未知回归函数 $m(x)$. 具有连续二阶导数的分段多项式 (立方样条) 常用作这样的逼近. 为明确起见，我们仅考虑立方样条逼近，但所介绍的思想也可以应用到其他阶的样条. 设 $t_1, \cdots, t_J$ 是一个固定的节序列，使得 $-\infty < t_1 < \cdots < t_J < \infty$. 那么，立方样条函数是二次连续可微函数 $s(x)$，使得 $s(x)$ 对区间 $(-\infty, t_1]$, $(t_1, t_2], \cdots, (t_{J-1}, t_J], (t_J, \infty)$ 的每一个限制为立方多项式. 所有立方样条函数组成 $J+4$ 维线性空间. 有两个这种线性空间的立方样条基：幂基和 B 样条基. 幂基的形式为

$$1, x, x^2, x^3, (x-t_1)_+^3, \cdots, (x-t_J)_+^3. \tag{1.2.12}$$

通常，B 样条基在计算上更稳定，因为基函数间的多重相关性小；而幂样条基具有这样的优越性：删除基函数 $(x-t_j)_+^3$ 与删除节 t_j 一样.

设 $B(x) = (B_1(x), \cdots, B_{J+4}(x))^{\mathrm{T}}$ 是一个立方样条基. 那么，立方样条函数可表示为

$$s(x) = \gamma^{\mathrm{T}} B(x),$$

其中 $\gamma = (\gamma_1, \cdots, \gamma_{J+4})^{\mathrm{T}}$ 为样条系数. $m(x)$ 可以用 $s(x)$ 逼近. 定义目标函数

$$\sum_{i=1}^{n} \left[Y_i - \gamma^{\mathrm{T}} B(X_i)\right]^2. \tag{1.2.13}$$

记 $\hat{\gamma} = (\hat{\gamma}_1, \cdots, \hat{\gamma}_{J+4})^{\mathrm{T}}$ 为极小化 (1.2.13) 的解, 那么回归函数 $m(x)$ 的样条最小二乘估计为

$$\hat{m}_{\mathrm{B}}(x) = \hat{\gamma}^{\mathrm{T}} B(x).$$

上述样条方法对节点数 J 和节点位置的选择是敏感的. 通常把节点放在曲率变化大的位置上. 我们使用节点删除法选取节点, 下面介绍它的基本思想. 设 $t_1, \cdots, t_J$ 是在节点选择过程中可以被删除的初始节点. 节点数 J 通常取为 $J = [n/2]$ 或 $J = [n/3]$. 通常把这些节点放在 X 变量的次序统计量上, 即 $t_j = X_{(2j)}$, $j = 1, \cdots, [n/2]$. 设相应的幂基如同 (1.2.12). 删除具有最小绝对 t 值 $|\hat{\gamma}_j|/\mathrm{SE}(\hat{\gamma}_j)(1 \leqslant j \leqslant J)$ 的第 $j_0(1 \leqslant j_0 \leqslant J)$ 个节点, 其中 $\mathrm{SE}(\hat{\gamma}_j)$ 为 $\hat{\gamma}_j$ 的标准误. 重复上述删除过程 (每步删除一个节点). 我们获得一列由 $j(0 \leqslant j \leqslant J)$ 标示的模型: 第 j 个模型包含具有残差平方和 RRS_j 的 $J+4-j$ 个自由参数. 极小化修正的 Mallows C_p 准则 (Mallows, 1973):

$$C_j = \mathrm{RSS}_j + \alpha(J + 4 - j)\hat{\sigma}^2, \tag{1.2.14}$$

可选出模型 $\hat{j}$, 其中 $\hat{\sigma}$ 是在第 0 个模型 (完整模型) 上估计的标准误, α 是对太复杂模型惩罚的光滑参数. 这就完成了节点的选择过程. 第 $\hat{j}$ 个模型涉及的节点用作最后的样条建模. 这个思想常被 Stone 和他的合作者所使用, 如 Koopergerg 和 Stone(1991), Koopergerg 等 (1995a, 1995b).

节点个数 J 在拟合数据和非参数函数估计的光滑程度之间起平衡作用. 随着节点个数的增加, 样条估计的方差随之增大, 而偏差变小, 此时估计会过分地拟合数据. 另一方面, 随着节点个数的减少, 样条估计的方差随之变小, 但是偏差变大, 此时估计出的函数会充分地光滑. 因此, 在实际应用中, 选取合适的节点个数往往是很重要的.

2. *光滑样条*

节点的自动选择的另一个方法是光滑样条. 下面介绍这种方法. 设 $\mathcal{G}_2$ 是 $\mathbf{R}$ 上所有连续可微、二次可积且具有二阶导数的函数构成的类. 对 $\mathcal{G}_2$ 中的任何函数 m, 定义准则

$$C_\lambda(m) = \frac{1}{n} \sum_{i=1}^{n} [Y_i - m(X_i)]^2 + \lambda \int [m''(x)]^2 \mathrm{d}x, \tag{1.2.15}$$

其中 λ 是一个正的常数, 称为光滑参数. (1.2.15) 式右边包含两项, 第一项提供了 m 对数据拟合程度的度量, 第二项用来度量 m 的光滑度. 常数 λ 反映了拟合的相对重要性和 m 的光滑性; 小的 λ 表示拟合更重要, 反之大的 λ 表示函数的光滑比拟合更重要.

极小化 $C_\lambda(m)$ 得到 $\hat{m}_{\mathrm{SS}}$, 称它为 m 的光滑样条估计量. 尤其, $\hat{m}_{\mathrm{SS}}$ 具有下列性质: 一是它在 $X_1,\cdots,X_n$ 处有节; 二是它在每一个区间 $[X_{i-1},X_i]$ 上是一个立方多项式, $i=1,\cdots,n$; 三是它具有二阶连续的导数.

考虑 $\lambda=0$ 和 $\lambda=\infty$ 两种极端情况是有价值的. 显然, 当 $\lambda=0$ 时, 相应的解是一个内插函数 $m(X_i)=Y_i$, 而 $\lambda=\infty$ 时导致一个线性回归 $m(x)=\alpha+\beta x$. 当 λ 在 0 到 ∞ 范围内变化时, $m(x)$ 的估计在最复杂的模型 (内插) 和最简单的模型 (线性模型) 之间变化. 因此, 光滑样条方法的模型复杂性由光滑参数 λ 有效地控制.

众所周知, 对 (1.2.15) 式中 $C_\lambda(m)$ 的极小化的解是区间 $[X_{(1)},X_{(n)}]$ 上的立方样条. 此解在这个数据范围内也是唯一的. 此外, 容易推出 $\hat{m}_{\mathrm{SS}}(x)$ 关于响应变量是线性的, 即

$$\hat{m}_{\mathrm{SS}}(x)=\frac{1}{n}\sum_{i=1}^{n}W_{ni}^{\mathrm{S}}(x)Y_i, \tag{1.2.16}$$

其中权 $W_{ni}^{\mathrm{S}}(x)=W_{ni}(x;X_1,\cdots,X_n)$ 不依赖于响应 $\{Y_i\}$. 见文献 (Härdle, 1990) 的 59–60 页. Silverman(1984) 指出: 光滑样条本质上是具有变带宽的局部核平均. 对远离边界的 X_i 以及对大的 n 和相对小的 λ,

$$W_{ni}^{\mathrm{S}}(x)\approx[f(X_i)h(X_i)]^{-1}K\left(\frac{X_i-x}{h(X_i)}\right), \tag{1.2.17}$$

其中 $h(X_i)=\{\lambda/[nf(X_i)]\}^{1/4}$,

$$K(t)=0.5\exp(-|t|/\sqrt{2})\sin(|t|/\sqrt{2}+\pi/4).$$

正如 Messer(1991) 所说的, 这个逼近也可以用来计算光滑样条估计量 $\hat{m}_{\mathrm{SS}}(x)$ 的均值和方差. 近似公式 (1.2.17) 提供了从直观上理解光滑样条在点 x 的局部邻域内指定权的形式.

多项式样条与光滑样条估计方法都是用样条函数来估计未知的非参数回归函数, 其主要区别在于光滑样条估计方法保守地选择较多数目的节点, 然后再通过惩罚系数防止估计过分地拟合数据. 而多项式样条估计方法直接通过节点来平衡拟合数据和估计函数的光滑度. 因此, 相对于光滑样条方法, 多项式样条估计在一般情况下所需要的节点个数较少, 从而待估的参数也较少.

1.2.4　权函数法与其他估计

前面几节介绍了三种光滑方法：核光滑，局部多项式光滑，样条光滑．由这些估计方法所构造的估计量都可以用某个概率权函数 $\{W_{ni}(x), 1 \leqslant i \leqslant n\}$ 表示．所谓概率权函数是指 $\{W_{ni}(x), 1 \leqslant i \leqslant n\}$ 满足：

$$W_{ni}(x) \geqslant 0,\ 1 \leqslant i \leqslant n, \quad \sum_{i=1}^{n} W_{ni}(x) = 1.$$

对给定的概率权函数 $\{W_{ni}(x), 1 \leqslant i \leqslant n\}$，定义回归函数 $m(x)$ 的估计

$$\hat{m}_{\mathrm{W}}(x) = \sum_{i=1}^{n} W_{ni}(x) Y_i, \tag{1.2.18}$$

并称 $\hat{m}_{\mathrm{W}}(x)$ 为 $m(x)$ 的一个权函数估计．

从 (1.2.18) 式可以看出：一个权函数估计完全由给定的权 $\{W_{ni}(x), 1 \leqslant i \leqslant n\}$ 所确定，而权函数的分布只与 X 的分布有关．样本 (X_i, Y_i) 对 $\hat{m}_{\mathrm{W}}(x)$ 的贡献，除本身值外，还取决于权 $W_{ni}(x)$．因而，权 $W_{ni}(x)$ 表示在估计 $m(x)$ 时，样本 (X_i, Y_i) 所起的作用的 "大小"．由上述定义可知，(1.2.2) 和 (1.2.10) 式所确定的权函数都是概率权函数．

除上述介绍的几种回归函数的估计方法之外，还有其他估计方法，例如，k-近邻法、小波光滑法、傅里叶级数光滑法等．k-近邻估计是用最靠近 x 的 k 个观察值所对应的 k 个响应变量的加权平均．k-近邻权序列由 Loftsgaarden 和 Quesenberry(1965) 在密度估计的相关领域引入．Donoho 和 Johnstone(1995) 将小波光滑成功地应用于非参数回归估计．关于小波估计的综述文献可参考文献 (Donoho et al., 1995)．傅里叶 (Fourier) 级数光滑与小波光滑属于同一类，文献中称为正交级数逼近法，用于函数的渐近表示、数据分解和曲线拟合等．上述光滑方法在相关文献中可以查到，这里不再一一赘述．

1.3　选择光滑参数

在上述非参数回归估计中，都是通过光滑参数控制建模的复杂性，如何决定光滑的程度是一个重要的问题．对于回归函数的核估计来说，太大的带宽往往得到过分光滑的曲线．如果带宽选的太小，得到的回归函数估计曲线就很不光滑．因此，人们把带宽也称为光滑参数．光滑样条估计量中的光滑参数也有类似的作用．如果光滑的目的是增加对噪声的光滑或是得到一个简单模型，那么可以使用带有主观选择光滑参数的 "过光滑 (oversmoothing)" 的曲线，另一方面，如果我们的兴趣

在于强调回归曲线的局部结构, 那么一条 "欠光滑 (undersmoothing)" 的曲线是适用的.

本节使用 h 表示任何一种非参数回归估计量的光滑参数, 包括核估计、局部多项式估计和样条估计等.

1.3.1 交叉验证

交叉验证 (cross-validation, CV) 方法是选择光滑参数的一个常用方法, 基本思想如下: 去掉第 i 个观测 (X_i, Y_i), 得到 "抛去一点" 的估计量

$$\hat{m}_h^{(-i)}(X_i) = \sum_{j \neq i} W_{nj}(X_i; h) Y_j, \tag{1.3.1}$$

其中 $\{W_{nj}(\cdot; h)\}$ 为概率权向量, 它没有使用第 i 个观测. 使用这个修正的估计量, 定义 CV 选择函数

$$\mathrm{CV}(h) = \frac{1}{n} \sum_{i=1}^{n} [Y_i - \hat{m}_h^{(-i)}(X_i)]^2. \tag{1.3.2}$$

那么, 自动光滑参数可定义为

$$\hat{h} = \arg\min_h \mathrm{CV}(h).$$

交叉验证方法使用了 Allen(1974) 的思想.

1.3.2 广义交叉验证

由于对每个检验的光滑参数统计量 (1.3.2) 的计算需要拟合 n 条曲线, 故计算量很大. 一个改良的形式是由 Wahba(1977) 与 Craven 和 Wahba(1979) 提出的广义交叉验证 (GCV). 下面介绍这个准则. 假设 $\hat{m}_h(x)$ 为 $m(x)$ 的任一个非参数回归估计, 那么可以把拟合值表示为

$$(\hat{m}_h(X_1), \cdots, \hat{m}_h(X_n))^{\mathrm{T}} = S_h \boldsymbol{Y}, \tag{1.3.3}$$

其中 S_h 为仅依赖于 X 变量的 $n \times n$ 帽子矩阵, $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^{\mathrm{T}}$. GCV 统计量定义为

$$\mathrm{GCV}(h) = \frac{n^{-1} \sum_{i=1}^{n} [Y_i - \hat{m}_h(X_i)]^2}{[n^{-1} \mathrm{tr}(I - S_h)]^2}. \tag{1.3.4}$$

极小化 GCV(h) 即可得到所需要的光滑参数. Wahba 和 Wang(1990) 给出了上述方法的描述.

1.4 经验似然

经验似然方法由 Thomas 和 Grunkemeier(1975) 首次提出, 并将该方法应用于生存概率的区间估计之中. 随后, Owen(1988) 用经验似然方法研究了总体均值的置信域的构造问题. 经验似然是一种非参数统计推断方法. 与传统的统计方法相比, 该方法具有很好的性质. 比如, 用经验似然方法构造置信域时, 不需要构造枢轴统计量, 所得置信域具有域不变性, 且其形状完全由数据自行确定. 因此, 经验似然方法引起了许多统计学家的关注, 并得到了长足发展. 下面介绍经验似然方法的统计思想和方法.

1.4.1 经验似然的思想

设 $X_1, \cdots, X_n$ 是来自总体 $X \in \mathbf{R}^p (p \geqslant 1)$ 的相互独立的随机样本, 且具有共同的分布 F_0. 样本的经验分布函数为

$$F_n(A) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in A),$$

其中 $I(X_i \in A)$ 表示任一事件 $A \subseteq \mathbf{R}^p$ 的示性函数. 那么关于 $F(x)$ 的非参数似然函数可定义为

$$L(F) = \prod_{i=1}^{n} F(\{X_i\}),$$

其中 $F(\{X_i\})$ 是在来自 F 的样本上得到值 X_i 的概率. 特别地, 当 $p = 1$ 时, $F(\{X_i\}) = P\{X = X_i\}, i = 1, \cdots, n$. Owen(2001) 证明了下述结果: 如果 $F \neq F_n$, 那么 $L(F) < L(F_n)$. 这个结果表明 F_n 使得上式达到极大, 即 F_n 为 F 的非参数极大似然估计. 因此, 非参数似然比可定义为

$$R(F) = \frac{L(F)}{L(F_n)}.$$

称 $R(F)$ 为 F 的经验似然比函数. 当 X_i 之间没有结点时, $R(F)$ 具有形式

$$R(F) = \prod_{i=1}^{n} np_i, \quad p_i = F(\{X_i\}),$$

这是因为 $L(F_n) = (1/n)^n$. Owen(1988) 证明了该公式在数据中有结时仍适合, 这只需修正 $\sum_{j: X_j = X_i} p_j = F(\{X_i\})$ 即可. 换句话说, 数据之间的结并不影响似然比的重新表示.

假设感兴趣的中心点是 $T(F_0)$, 其中 $T(\cdot)$ 是一个函数. $T(F_0)$ 的非参数极大似然估计是 $T(F_n)$. Owen(1988,1990) 在一些合理的条件下证明了集合

$$\{T(F)|R(F) \geqslant r\}$$

可以用来作为 $T(F_0)$ 的置信域. 为使这样的域有很好的性质, 需要限制 $F \ll F_n$.

Owen (1988,1990) 在相当一般的条件下证明了 $-2\log(R(F))$ 以分布收敛到自由度为 q 的 χ_q^2 变量, 其中 q 为待估参数的个数. 称这个渐近性质为非参数版本的 Wilks 定理, 它可以被用来构造待估参数的置信域.

1.4.2 参数的经验似然

下面用类似于参数似然的思想解释上述经验似然. 对某个函数 T, 假设我们对参数 $\theta = T(F)$ 感兴趣, 其中 F 是集 $\mathcal{F}$ 的元素. 在一些情况下, 我们可以取 $\mathcal{F}$ 为 $\mathbf{R}$ 上所有分布的集合. 往往使用分布的最小集. 定义截面似然比函数

$$\mathcal{R}(\theta) = \sup\{R(F)|T(F) = \theta, F \in \mathcal{F}\}.$$

对某个临界值 r_0, 当 $\mathcal{R}(\theta_0) < r_0$ 时, 经验似然检验拒绝 $\mathcal{R}(\theta_0) < r_0$. θ 的经验似然置信域为

$$\{\theta|\mathcal{R}(\theta) \geqslant r_0\}.$$

可以证明, $\mathcal{R}(\theta)$ 具有非参数版本的 Wilks 定理. 因此, 临界值 r_0 可以用 χ^2 分布的分位点得到.

假设 X 具有均值 $\mu_0 = (\mu_{10}, \cdots, \mu_{p0}) \in \mathbf{R}^p$ 和满秩的方差 V_0. 为构造 μ_0 的经验似然置信域, Owen(1988) 定义了如下经验似然比函数

$$\mathcal{R}(\mu) = \max\left\{\prod_{i=1}^n np_i \middle| p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i X_i = \mu\right\}.$$

它要求在满足约束条件 $\sum\limits_{i=1}^n p_i X_i = \mu$ 的情况下使得非参数似然比达到极大. 而兴趣参数 μ 通过约束条件引入到似然比中, 从而得到 μ 的非参数似然比函数, 并且利用它作区间估计、假设检验和其他统计推断.

Owen(1990) 证明了 $-2\log\mathcal{R}(\mu_0) \xrightarrow{D} \chi_p^2$. 因此, μ_0 的近似 $1-\alpha$ 置信域为

$$C_{\mu_0} = \{\mu| - 2\log\mathcal{R}(\mu) \leqslant \chi_p^2(1-\alpha)\},$$

其中 $\chi_p^2(1-\alpha)$ 满足 $P\{\chi_p^2 \leqslant \chi_p^2(1-\alpha)\} = 1-\alpha$, 称之为自由度为 p 的 χ^2 分布的 $1-\alpha$ 分位点.

1.4.3　估计方程与经验似然

下面讨论经验似然和估计方程之间的联系. 对一个随机变量 $X \in \mathbf{R}^d$ 与一个参数 $\theta \in \mathbf{R}^p$ 以及一个向量函数 $m(X,\theta) \in \mathbf{R}^s$, 假设

$$E[m(X,\theta)] = 0. \tag{1.4.1}$$

通常设 $p = s$, 在关于 $m(X,\theta)$ 和 F 的一些条件下, 方程 (1.4.1) 有唯一解 θ. 在这种恰好确定的情况下, 通过解关于 $\hat{\theta}$ 的方程

$$\frac{1}{n}\sum_{i=1}^{n} m(X,\hat{\theta}) = 0 \tag{1.4.2}$$

可以估计真值 θ_0. 例如, 若取 $m(X,\theta) = X - \theta$, 那么方程 (1.4.2) 给出 $\hat{\theta} = \bar{X}$. 方程 (1.4.2) 是熟知的估计方程, 称 $m(X,\theta)$ 为估计函数. 大多数极大似然估计量都是通过估计方程来定义的.

当 $s < p$ 时, 方程 (1.4.1) 和 (1.4.2) 都有 θ 的 $p - s$ 维解集. θ 的某些函数可恰好由这个解集确定. 当 $s > p$ 时, 即使 (1.4.1) 式对真值 F_0 成立, 但对非参数极大似然估计 $\hat{F}$ 通常不成立. 广义矩方法可得到接近于方程 (1.4.2) 的解的一个值 $\hat{\theta}$. 经验似然方法也可以解决这个问题, 它与估计方程有密切的联系. θ 的经验似然比函数可定义为

$$\mathcal{R}(\theta) = \max\left\{\prod_{i=1}^{n} np_i \middle| p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i m(X_i,\theta) = 0\right\}.$$

Owen(2001) 在适当的条件下证明了 $-2\log\mathcal{R}(\theta_0) \xrightarrow{D} \chi_p^2$, 其中 θ_0 满足 $E[m(X,\theta)] = 0$.

1.4.4　主要成果与文献注记

Owen(1991) 将经验似然方法应用到了线性模型的统计推断之中. Chen(1993, 1994) 深入研究了线性模型中回归系数的经验似然置信区间的精度问题. Kolaczyk (1994) 与 Chen 和 Cui(2003) 及 Xue 等 (2011) 进一步将经验似然推广到广义线性模型的研究中. Owen(2001) 在他的专著中对经验似然方法作了详细阐述.

近十多年来, 经验似然方法已经应用到非参数和半参数模型的统计推断之中. Hall 和 Owen(1993) 研究了密度函数的经验似然置信带的构造问题. Chen 和 Qin (2000) 利用经验似然方法构造了回归函数的估计. Wang 和 Jing(1999) 与 Shi 和 Lau(2000) 分别考虑了部分线性模型, 构造了模型中未知参数的大样本置信域. Xue 和 Zhu(2005,2006) 利用经验似然方法研究了单指标模型和部分线性单指标模型中指标系数的置信域构造问题. Zhu 和 Xue(2006) 在研究部分线性单指标模型时提出了一种 “纠偏” 的经验似然方法, 不仅可以使所构造的纠偏的经验对数似然比统计

量渐近于标准 χ^2 分布, 而且还放宽了带宽的选取范围, 避免了对非参数函数的欠光滑 (undersmoothing). You 和 Zhou(2006) 与 Huang 和 Zhang(2009) 把经验似然方法应用于半参数变系数部分线性模型中参数和非参数分量的统计推断之中. 其他方面的工作参看文献 (Chen and Hall, 1993; Qin, 1993; Chen and Sitter, 1999; Qin and Lawless, 1994; Peng, 2004) 等.

Kitamura(1997) 使用分组技巧把经验似然方法从独立同分布的数据推广到相依数据. Mykland(1995,1999) 在鞅差序列情形下利用经验似然方法处理相依数据, 并证明了经验似然在相依数据中的 Bartlett 纠偏性. Dicicio 等 (1991) 证明了经验似然具有 Bartlett 纠偏性, 可以把覆盖误差由 $O(n^{-1})$ 减少到 $O(n^{-2})$, 其中 n 表示样本容量. 自从这一结果提出以来, 已经有许多作者对 Bartlett 纠偏性进行了更深入的研究, 相关的结果可见文献 (Thomas and Joseph, 1989; Jing and Wood, 1996; Corcoran, 1998; Diciccio et al., 2000). Baggerly(1998) 证明了在所有经验 Cressie-Read 似然族中只有经验似然具有 Bartlett 纠偏性.

近年来, 一些学者在用经验似然方法对复杂数据的统计分析方面做了大量工作, 其中所考虑的数据包括: 删失数据、缺失数据、测量误差数据、纵向数据等. Wang 和 Jing(2001) 与 Li 和 Wang(2003) 在随机删失下发展了生存分布一类泛函以及随机删失线性和部分线性模型的统计推断. 对于缺失数据下的统计模型, 许多学者在线性回归借补、非参数回归借补和半参数回归借补下发展了经验似然方法. 在这方面的主要研究成果可参看文献 (Wang and Rao, 2001, 2002b, 2002c; Wang et al., 2004; Qin and Zhang, 2007; Wang and Sun, 2007; Xue, 2009a, 2009b; Xue and Xue, 2011; Xue et al., 2011). 对于在测量误差模型中的应用, Wang 和 Rao(2002a) 与 Stute 等 (2007) 在核实数据的帮助下, 分别把经验似然方法推广到了协变量带有测量误差的线性模型和非线性模型之中, 定义了估计的经验似然和调整的经验似然, 并证明了它们的渐近 χ^2 性. 此外, Cui 和 Chen(2003) 与 Cui 和 Kong(2006) 在可加的误差模型结构下, 分别构造了线性及部分线性模型中未知参数的经验似然置信域. 其他相关的工作有 (Cui and Li, 1998; Cui and Zhu, 2003). 近几年有一些关于在纵向数据情形下进行经验似然统计推断的研究成果. Xue 和 Zhu(2007a) 利用经验似然的方法研究了纵向数据下变系数模型, 提出了新的纠偏方法, 构造了均值修正的经验似然比和残差调整的经验似然比, 从而可以用来构造模型中系数函数的置信带, 发展了经验似然的理论和方法. Xue 和 Zhu(2007b) 构造了纵向数据部分线性模型中参数分量的置信域和非参数分量的逐点置信区间. Xue(2010) 将经验似然方法和局部多项式光滑相结合, 在类数据下构造了回归函数及其导函数的置信带. You 等 (2006) 考虑了纵向数据下半参数部分线性模型, 提出了分组的经验对数似然比统计量. Li 等 (2010) 用经验似然方法研究了纵向数据的部分线性单指标模型, 构造了兴趣参数的置信域.

1.5 惩罚方法

变量选择是现代统计分析中的重要课题. 传统的变量选择方法又称为子集选择方法 (subset selection), 即对协变量集合的所有子集进行比较分析, 通过假设检验或诸如 AIC 和 BIC 信息准则, 从中选出一个最优的子集来拟合回归模型. 这些方法的一个共性是根据已有样本用一个准则选择出变量子集, 然后再基于这一样本来估计回归系数. 但是, 由于真实的相关变量往往是不知道的, 而选择变量的过程也会产生一定的偏差. 这就很难评价最终模型中回归系数估计的精度. 另外, 当协变量的维数 p 较大时, 对所有的 2^p-1 个子集进行假设检验, 计算量相当大, 在实际应用中也难于实现. 因此, 利用惩罚方法进行变量选择越来越受到统计学家的重视. 惩罚方法的基本思想就是在进行参数估计的同时, 把较小的估计系数压缩为 0, 从而达到变量选择的目的. 该方法可以同时进行参数估计和变量选择, 从而大大减少了计算量, 并且克服了传统变量选择的不稳定性. 目前, 已有大量文献利用惩罚方法研究变量选择问题.

1.5.1 惩罚函数

假设 p 维协变量 X_i 中有一些分量不是统计显著的, 模型参数 β 是 p 维的, $l(\beta)$ 为给定的二次损失函数. 惩罚最小二乘定义为

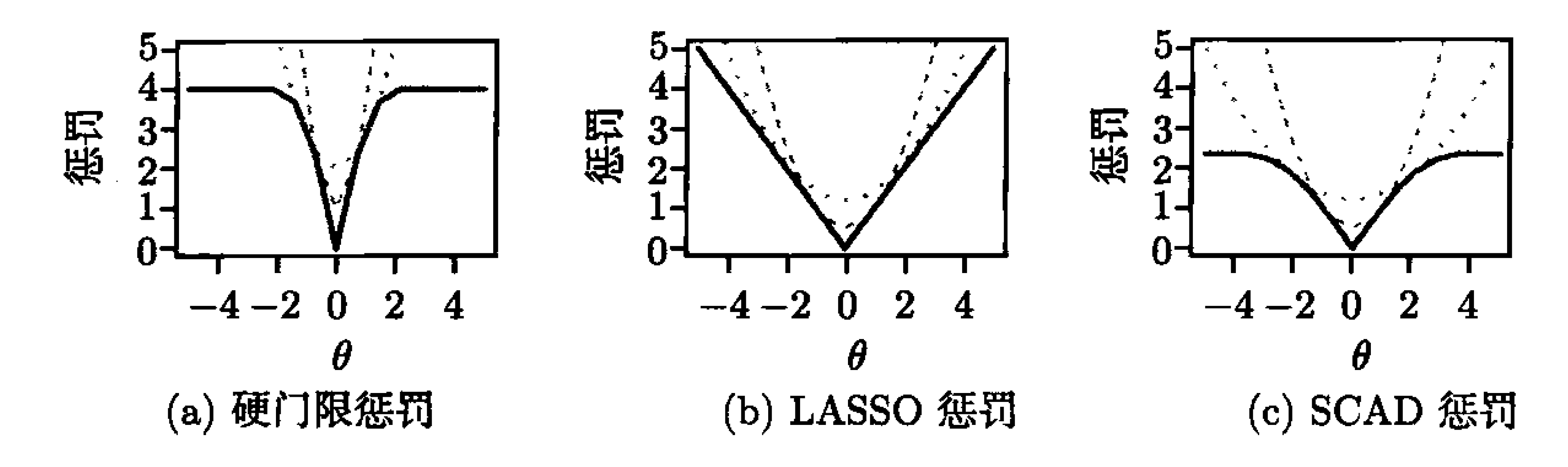
$$\mathcal{L}(\beta)=l(\beta)+n\sum_{j=1}^{p}\lambda_j p_j(|\beta_j|), \tag{1.5.1}$$

其中 $p_j(\cdot)$ 为惩罚函数, λ_j 为调整参数, 它是用来控制模型的复杂度, 并可以用诸如 CV 或 GCV 的数据驱动方法来选择. 通过极小化 $\mathcal{L}(\beta)$, 一些系数被估为 0, 那么删除相应的变量, 否则保留该变量. 这个方法可以同时选择变量和估计系数. 所得到的估计量称为惩罚最小二乘估计. 惩罚函数 $p_j(\cdot)$ 和调整参数 λ_j 不必要对所有的 j 都相同. 这就允许我们通过使用不同的惩罚函数或取 λ_j 的不同值来合并未知系数的先验信息. 例如, 我们希望保留参数模型的显著性预测, 因此将不需要惩罚他们的相应的参数. 为表达方便, 假定惩罚函数和调整参数对所有系数都相同, 并用 $p_\lambda(\cdot)$ 表示 $\lambda p(\cdot)$. 扩展到不同的门限函数的情况不牵涉任何额外的难度.

许多惩罚函数已经被用到各种参数模型的惩罚最小二乘和惩罚似然之中. 例如, $L_q(q\geqslant 0)$ 惩罚类: L_0 惩罚 $p_\lambda(|\theta|)=0.5\lambda^2 I(|\theta|\neq 0)$ 相当于熵惩罚; L_1 惩罚 $p_\lambda(|\theta|)=\lambda|\theta|$ 导致 Tibshirani(1996) 提出的 LASSO 惩罚, 见图 1.5.1(b); L_2 惩罚 $p_\lambda(|\theta|)=\lambda|\theta|^2$ 引起岭回归; 桥回归 (Frank and Friedman, 1993) 相当于 $p_\lambda(|\theta|)=\lambda|\theta|^q(0<q<1)$. Antoniadis(1997) 考虑了硬门限惩罚函数

$$p_\lambda(|\theta|)=\lambda^2-(|\theta|-\lambda)^2 I(|\theta|<\lambda), \tag{1.5.2}$$

见图 1.5.1(a). 该图说明了通过对参数的惩罚可以将小的估计参数压缩为 0.



(a) 硬门限惩罚 (b) LASSO 惩罚 (c) SCAD 惩罚

图 1.5.1 三个惩罚函数和它们的二次逼近, 其中 (a) $\lambda=2$, (b)、(c) $\lambda=1$

在惩罚最小二乘中, 一个关键的问题是如何选择惩罚函数. Antoniadis 和 Fan (2001) 与 Fan 和 Li(2001) 对如何选择惩罚函数提出了深刻的见解. 他们主张一个好的惩罚函数将导致具有三个性质的估计量.

(1) **无偏性**: 当真参数很大时, 所得到的估计量是几乎无偏的, 以避免不必要的建模偏差.

(2) **稀疏性**: 所得到的估计量是一个门限值, 自动将小的估计系数设置为 0, 以减少模型复杂性.

(3) **连续型**: 所得到的估计量在数据点处是连续的, 以避免模型预测中的不稳定.

Fan 和 Li(2001) 讨论了具有上述性质的惩罚函数所满足的条件, 他们得到如下结论: 大的真参数估计具有无偏性的充要条件是对大的 $|\theta|$ 有 $p'_\lambda(|\theta|))=0$; 具有稀疏性的充分条件是 $\min_\theta\{|\theta|+p'_\lambda(|\theta|)\}>0$; 具有连续性的充要条件是 $\arg\min_\theta\{|\theta|+p'_\lambda(|\theta|)\}=0$. 因此, 满足稀疏性和连续性的惩罚函数一定在原点处奇异, 见图 1.5.1.

L_q 惩罚方法所得到的估计量都不能同时满足上述三个性质. LASSO 惩罚方法所得到的估计量是有偏的, 而硬门限惩罚方法所得到的估计量是不连续的. 为了克服现有方法的缺陷, Fan 和 Li(2001) 提出了 SCAD 惩罚方法, 由该方法得到的估计量具有无偏性、稀疏性和连续型. SCAD 惩罚函数定义为

$$p_\lambda(|\theta|)=\begin{cases}\lambda|\theta|, & |\theta|\leqslant\lambda,\\ \dfrac{(a^2-1)\lambda^2-(|\theta|-a\lambda)^2}{2(a-1)}, & \lambda<|\theta|\leqslant a\lambda,\\ \dfrac{1}{2}(a+1)\lambda^2, & |\theta|>a\lambda.\end{cases} \tag{1.5.3}$$

见图 1.5.1(c). SCAD 惩罚函数包含两个未知常数. Fan 和 Li(2001) 从 Bayes 观点上建议取 $a=3.7$. λ 的选取可参见文献 (Fan and Li, 2001), 也可以用 CV 方法选取.

Fan 和 Li(2001) 证明了 SCAD 变量选择方法可以相合地识别出真实模型, 并且所得的正则估计是相合的, 并证明了非零系数的估计量与基于真实子模型所得的

估计量有相同的渐近分布. Fan 和 Peng(2004) 研究了参数维数发散的非凹惩罚似然, 在适当条件下证明了 SCAD 惩罚在这种情形下仍具有 oracle 性质. Zou(2006) 对 LASSO 惩罚函数进行了改进, 提出了自适应 LASSO 惩罚函数

$$p_\lambda(|\theta|) = \lambda|\theta|w, \tag{1.5.4}$$

其中 w 是已知权函数. 可取 $w = 1/|\hat{\theta}|^\lambda$, 其中 $\hat{\theta}$ 为 θ 的最小二乘估计. 一般取 $\lambda \in (0, 1)$ 或 $\lambda = 2$. Zou(2006) 证明了自适应 LASSO 惩罚估计也具有 oracle 性质.

1.5.2　惩罚估计与变量选择

考虑线性模型

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon, \tag{1.5.5}$$

其中 $\boldsymbol{Y}$ 是 $n \times 1$ 向量, $\boldsymbol{X}$ 是 $n \times p$ 矩阵. 正如传统的线性模型, 我们假定诸 Y_i 在给定设计矩阵下是相互独立的. 在线性模型中, 惩罚最小二乘与变量选择之间有很强的联系. 为了加深对各种变量选择方法的理解, 假定 (1.5.5) 式中的 $\boldsymbol{X}$ 的列是正交的. 通过极小化 $||\boldsymbol{Y} - \boldsymbol{X}\beta||^2$ 得到的最小二乘估计 $\hat{\beta}$ 等价于 $||\hat{\beta} - \beta||^2$ 的极小值点, 其中 $\hat{\beta} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$ 是普通的最小二乘估计.

记 $\boldsymbol{z} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$, 并设 $\hat{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$. 一个惩罚最小二乘是

$$\begin{aligned}\mathcal{L}_p(\beta) &= \frac{1}{2}||\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}||^2 + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \\ &= \frac{1}{2}||\boldsymbol{Y} - \hat{\boldsymbol{Y}}||^2 + \frac{1}{2}\sum_{j=1}^{p}(z_j - \beta_j)^2 + \sum_{j=1}^{p} p_\lambda(|\beta_j|).\end{aligned} \tag{1.5.6}$$

因此, 极小化 $\mathcal{L}_p(\beta)$ 等价于逐分量极小化. 这就启发我们考虑惩罚最小二乘问题

$$\frac{1}{2}(z - \theta)^2 + p_\lambda(|\theta|). \tag{1.5.7}$$

若取硬门限惩罚函数 (1.5.2), 则得到硬门限值

$$\hat{\theta}_{\mathrm{H}} = zI(|z| > \lambda). \tag{1.5.8}$$

换句话说, 极小化 $\mathcal{L}_p(\beta)$ 的解仅仅是 $z_j I(|z_j| > \lambda)$, 这就与最好佳集选择和逐步删除相一致. 显然, $\hat{\theta}_{\mathrm{H}}$ 是不连续的, 见图 1.5.2(a). LASSO 惩罚函数 $p_\lambda(|\theta|) = \lambda|\theta|$ 产生一个软门限值

$$\hat{\theta}_{\mathrm{L}} = \mathrm{sgn}(z)(|z| - \lambda)_+. \tag{1.5.9}$$

可以证明 $\hat{\theta}_{\rm L}$ 不具有无偏性, 但它是连续的, 见图 1.5.2(b). SCAD 惩罚函数 (1.5.3) 可获得解

$$\hat{\theta}_{\rm S}=\begin{cases}\operatorname{sgn}(z)(|z|-\lambda)_+, & |z|\leqslant 2\lambda,\\ \dfrac{(a-1)z-\operatorname{sgn}(z)a\lambda}{a-2}, & 2\lambda<|z|\leqslant a\lambda,\\ z, & |z|>a\lambda.\end{cases}\tag{1.5.10}$$

参见图 1.5.2(c). (1.5.10) 式的门限值包含两个未知参数 λ 和 a. 在实际中, 我们可以使用 CV 或 GCV 准则在二维格子点上搜索最佳数对 (λ,a). 然而, 实施这样的计算可能是很费时的. Fan 和 Li(2001) 使用 Bayes 风险分析方法得到了 λ 和 a 的理想值, 即 $\lambda=\sqrt{a\log(d)}$, $a\approx 3.7$, $d<100$. 容易验证 $\hat{\theta}_{\rm S}$ 具有无偏性、稀疏性和连续型.

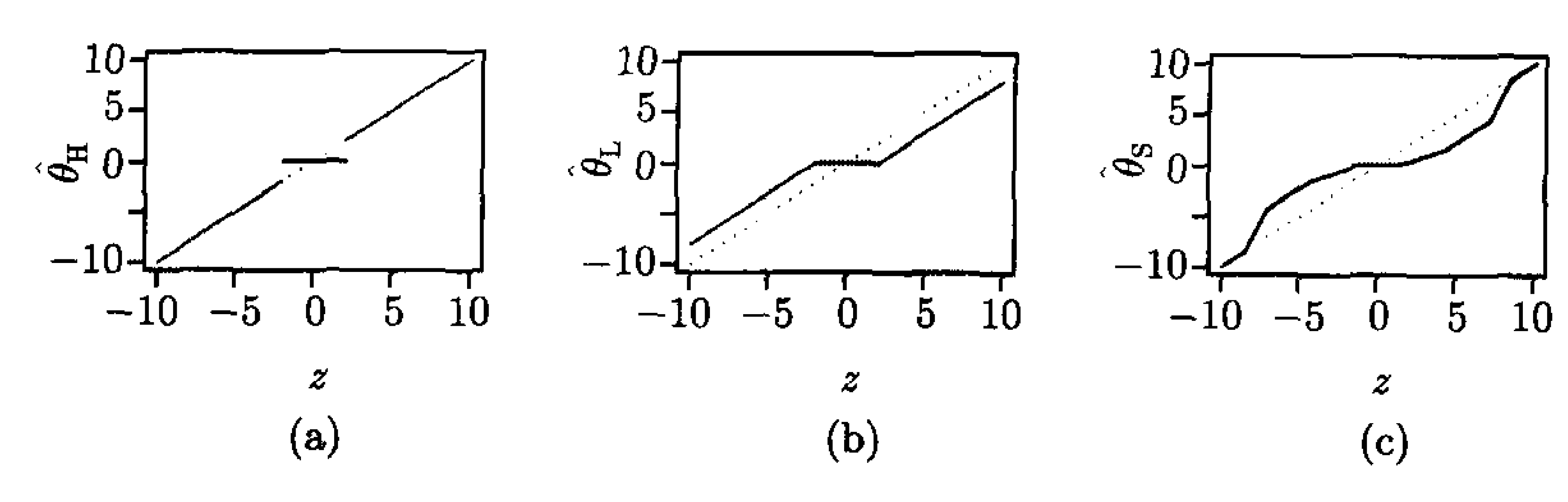


图 1.5.2 门限函数图: (a) 硬门限, (b) LASSO 门限, (c) SCAD 门限, 其中 $\lambda=2$, $a=3.7$

1.5.3 主要成果与文献注记

由于 SCAD 变量选择方法具有很好的性质, 目前该方法已被广泛地使用于半参数模型的变量选择之中, 例如, 部分线性模型、单指标模型、变系数模型等. Li 和 Liang(2008) 将 SCAD 惩罚应用到了半参数模型的变量选择的研究之中, 对参数部分, 采用 SCAD 惩罚提出了一系列变量选择方法, 证明了所得的惩罚估计具有 oracle 性质; 对非参数部分, 提出了半参数广义似然比检验, 通过检验选出非参数部分中重要的变量, 同时研究了在零假设下所提出的检验统计量的渐近性质. Fan 和 Li(2004) 考虑了纵向数据下部分线性模型的变量选择, 首先假定参数分量已知, 用局部多项式方法估计非参数部分, 然后采用 SCAD 惩罚构造惩罚函数, 通过惩罚函数获得 SCAD 惩罚估计, 并证明了惩罚估计具有 oracle 性质. Xie 和 Huang(2009) 讨论了参数维数发散的部分线性模型的变量选择问题, 先用多项式样条方法估计非参数函数, 然后采用 SCAD 惩罚得到惩罚估计, 并研究了所得估计量的收敛速度和渐近性质. Zhu 和 Zhu(2009) 考虑了高维单指标模型的变量选择, 他们采用 SCAD 惩罚方法估计参数部分, 结果表明所得估计具有 oracle 性质. Liang 和 Li(2009) 研究了协变量带有测量误差的部分线性模型的变量选择问题, 提出了两种变量选择方法, 一种是采用惩罚最小二乘方法, 另一种是惩罚分位数回归. 他们采用 SCAD 惩

罚去构造惩罚最小二乘估计, 证明了所得估计量具有了 oracle 性质. Koenker(2005) 采用 SCAD 惩罚构造惩罚分位数函数, 所得的估计的 oracle 性质仍成立. 除了上述提到的文献外, 关于变量选择的文献还有 (Cai et al., 2005; Zou and Hastie, 2005; Wang et al. 2007; Zou et al., 2007; Wang et al., 2008; Zou and Li, 2008; Zou and Yuan, 2008; Fan and Lv, 2009; Zou and Zhang, 2009) 等.

1.6　复杂数据简介

在实际问题中, 人们经常会遇到一些复杂数据, 如纵向数据、缺失数据、删失数据、测量误差数据等. 在统计分析中, 随着数据的不同, 处理方法也有较大差异. 本节简单介绍上述四种常见的复杂数据.

1.6.1　纵向数据

纵向数据 (longitudinal data) 是对同一组受试个体在不同时间点上重复观测的数据 (Diggle et al., 2002). 这类数据在生物医学、社会科学、计量经济学等众多领域普遍存在. 考虑含有 n 个个体的样本, 对第 i 个个体在时间点 $t = t_{ij}(i = 1,\cdots,n, j = 1,\cdots,m_i)$ 处对响应变量 $Y_i(t)$ 和协变量 $X_i(t)$ 进行观测, 其中 m_i 表示对第 i 个个体总的观测次数. 那么可得纵向数据 $\{(t_{ij}, X_i(t_{ij}), Y_i(t_{ij})), 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m_i\}$. 纵向数据与截面数据 (cross-sectional data) 不同, 截面数据是指仅在某一个时间点对同一个体作的一次观测. 在纵向数据分析中, 研究的兴趣通常集中在评价时间 t 和协变量 X 对响应变量 Y 的效应, 其中协变量可以依赖于时间 t, 也可以不依赖于时间 t. 因此, 纵向数据样本也可表示为 $\{(t_{ij}, X_{ij}, Y_{ij}), 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m_i\}$. 虽然纵向数据组间是独立的, 但在同一个体内的不同观测往往具有相关性.

为了进一步理解纵向数据的特征, 下面用两个例子来加以说明.

例 1.6.1(癫痫病发作研究)　在对癫痫病发作的研究中, 59 个癫痫病患者参与了实验. 在实验期间, 将 59 个病人随机分为两组, 并采取两种不同的疗法, 其中 28 个病人服安慰剂, 31 个病人服抗痫药物. 每隔两星期对这些病人进行连续 4 次访问, 并记录他们之中癫痫病发作的人数. 在该项研究中, 响应变量为两星期期间癫痫病发作数量, 协变量分别为年龄的对数和基准癫痫病数 (除以 4 后取对数). 一个科学的问题是研究药物是否有助于降低癫痫病发作的频数. 对该问题的研究可参见文献 (Thall and Vail, 1990; Wang et al., 2005; Pang and Xue, 2012).

该实验是平衡纵向数据实验的一个例子, 也就是说, 每个个体在固定的时间点进行观测, 观测的次数相同.

例 1.6.2(CD4 细胞数研究)　自 1984 年以来, 国际艾滋病研究中心 (MACS)

记录了大约 5000 人的 CD4 细胞数的数据. Zeger 和 Diggle(1994) 将半参数模型应用到该数据集的统计分析. 免疫力缺乏病毒 (HIV) 通过降低一个人抵御感染的能力从而使人导致艾滋病. HIV 损害人体内一种称为 CD4 的细胞. 一个未感染的个体每毫升血液中大约含有 1100 个 CD4 的细胞. 当他感染了 HIV 后, 其 CD4 细胞数会随着时间逐步减少. 因此, 一个患者的 CD4 细胞数可以用来监测疾病的进程. 实验参加者每隔半年做一次 CD4 细胞数检查. MACS 研究的目的是刻画 CD4 细胞数损耗的平均时间趋势, 并针对每一个体估计其 CD4 细胞数损耗的时间趋势.

易见, 这个数据集是非平衡纵向数据, 每个个体的观测数目不等且观测是在完全任意的时间点进行的.

在以上两个例子中, 个体之间可以合理地假定为相互独立, 但每个个体内的多次重复观测很可能是相关的. 每种研究都可以归结为回归问题, 用来描述响应变量对协变量的依赖关系. 纵向数据对同一个个体进行重复观测的特点是它可以直接研究个体随时间的变化趋势.

对于纵向数据, 可以有几种不同的方法用于建模响应变量 Y 对协变量 X 的依赖关系.

第一种方法称为边缘模型 (marginal model). 所谓边缘是指响应变量的均值仅依赖于感兴趣的协变量, 而不依赖于任何随机效应. 边缘模型将响应变量对协变量的回归与组内相关性分离开来进行建模, 其研究的重点是回归系数的估计, 而把内相关视为讨厌参数. 这种方法类似于截面数据方法来建模边缘均值 $E(Y_{ij})$, 即假定 $E(Y_{ij})$ 为协变量 X_{ij} 的函数. 例如在 CD4 细胞数例子中, 总体平均 CD4 细胞数水平可以刻画为时间的函数. 既然重复观测之间不可能独立, 这种边缘分析方法也必须包括对相关形式的假定. 边缘回归系数的含义与截面数据中回归系数有相同的解释. 回归系数是基于总体的, 而不是基于个体的, 它表示协变量对总体平均的作用. 边缘模型方法具有分开建模均值和协方差的优点, 基于个体间的独立性, 即使当相关形式的假定不正确时, 仍能获得回归系数的有效推断.

第二种方法称为随机效应模型 (random effects model). 随机效应模型假定组内相关性是由于个体间回归系数的变化而产生的. 例如, 给定个体指定系数 (subject-specific coefficients)β_i, 建模 Y_{ij} 的条件期望为 $E(Y_{ij}|\beta_i) = X_{ij}^{\mathrm{T}}\beta_i$. 由于每个个体的重复观测数据太少而不能仅由 $\{(X_{ij}, Y_{ij})\}$ 估计 β_i, 我们进一步假定所有的 β_i 来自于具有均值 β 的某一分布. 如果我们记 $\beta_i = \beta + U_i$, 其中 β 是固定的, U_i 是 0 均值随机变量, 则基本的异方差假定可以根据潜在变量 (即随机效应)U_i 来重新描述. 也就是说, 存在潜在变量 U_i, 它对同一个个体的所有重复观测来说是相同的, 但在不同个体之间是变化的, 由此导致了组内相关性. 正因如此, 这种模型有时也称为潜在变量模型 (latent variable model). 在随机效应模型中, 回归系数表示协变量对个体的作用, 而不是对总体平均的作用. 当研究的目的是对个体进行推断, 而并非

总体平均时, 随机效应模型十分有用. 例如在例 1.6.2 中, 估计某一个体 CD4 细胞数损耗的时间进程就可以用随机效应来建模.

第三种方法称为转移模型 (transition model) 或马尔可夫模型 (Markov model). 这种模型是在给定过去响应变量的情况下来建模当前响应变量的条件均值, 即把条件均值 $\mathrm{E}(Y_{ij}|Y_{i(j-1)},\cdots,Y_{i1},X_{ij})$ 假定为协变量 X_{ij} 和过去响应变量 $Y_{i(j-1)},\cdots,Y_{i1}$ 的某一显示函数形式. 例子之一是有关二元数据的 logistic 回归模型

$$\log\frac{P(Y_{ij}=1|Y_{i(j-1)},\cdots,Y_{i1},X_{ij})}{1-P(Y_{ij}=1|Y_{i(j-1)},\cdots,Y_{i1},X_{ij})}=X_{ij}^{\mathrm{T}}\beta+\alpha Y_{i(j-1)}.$$

转移模型把 Y 关于 X 的依赖关系的假定和重复观测产生的相关的假定结合起来, 用一个简单的方程来表示. 由于先前观测 $Y_{i(j-1)},\cdots,Y_{i1}$ 影响了当前观测 Y_{ij}, 因而同一个体内重复观测 $Y_{i1},\cdots,Y_{im_i}$ 是相关的. 在转移模型中, 回归系数的解释依赖于个体的先前观测.

以上纵向数据分析的三种建模方法都既考虑了响应变量对协变量的依赖关系, 又考虑了个体内重复观测的相关性. 而对于截面数据, 仅需假定响应变量对协变量的依赖关系, 而无需考虑内相关性.

纵向数据分析问题可以划分为两大类:

(1) 研究的目的是建立响应变量对协变量的回归, 在此类问题中个体的数目远远大于每个个体重复观测的数目;

(2) 研究的主要兴趣是组内相关性, 在此类问题中个体的数目较小.

对于第 1 类问题, 研究者必须把主要的时间和精力用于正确地建模边缘均值, 而建模内相关性则是次要的. 由于独立个体的数目远大于组内重复观测的数目, 基于这一点可以得到回归参数估计量的稳健的方差估计, 即使内相关性错误地被指定, 这个稳健估计也可以用来获得回归参数的有效统计推断. 然而, 在第 2 类问题中, 为了得到有效推断, 必须较准确地同时建立均值和协方差模型.

1.6.2　缺失数据

在统计分析中, 数据缺失的现象是非常普遍的. 例如, 在临床实验时, 因各种原因存在缺失值, 如患者拒绝继续参加研究、治疗失败或成功、不良事件、患者搬家等; 在社会调查中，被调查者可能拒绝回答自己的收入情况; 在工业实验中, 由于各种原因部分实验结果记录不完整等. 这些不完整的数据给数据分析和应用带来了许多困难. 一些统计学家提出了不同的方法来处理缺失数据. 从缺失机制与方式上可将缺失数据分为三类, 即完全随机缺失、随机缺失和非随机缺失.

(1) 完全随机缺失 (missing completely at random, MCAR). 它指缺失现象是随机发生的, 即某一变量的缺失与非缺失之间不存在任何系统差异, 数据集中与其有关的其他变量, 在该变量的缺失与非缺失分组之间不存在任何系统差异, 可以说, 缺

失数据和观察都是总体的一个随机样本. 这类数据是缺失数据问题中最简单的一种. 在统计分析中, 若将缺失的记录删除, 估计结果不会有偏或偏差很小, 其检验功效与样本量的大小有关. 在对这类数据进行统计分析时所用到的模型都是可忽略模型 (ignorable model), 但实际数据分析中符合完全随机缺失的情况很少, 一般情况下数据缺失都是有其原因的.

(2) 随机缺失 (missing at random, MAR). 它是针对已获得的调查表中某一变量而出现的项目缺失. 一个变量是否随机缺失, 不是由这一变量本身所决定. 而是由数据集中与此变量有关的其他变量所决定. 如果在某变量的缺失与非缺失数据之间, 与该变量有关的其他变量间不存在差异, 那么这一变量的缺失为随机缺失.

(3) 非随机缺失 (missing not at random, MNAR). 如果数据不满足上述两种数据缺失方, 式则称为非随机缺失或非完全随机缺失. 这是缺失数据中处理起来最麻烦的一种.

近年来, 缺失数据问题在应用领域越来越引起人们的普遍关注. 在回归分析中, 数据缺失往往分为响应变量缺失和协变量缺失两种情况. 例如, 假定获得如下不完全数据

$$\{(Y_i, X_i, \delta_i),\ i = 1, \cdots, n\}, \tag{1.6.1}$$

其中协变量 X_i 都可观测到, 而响应变量 Y_i 缺失, 即若 Y_i 缺失, 则 $\delta_i = 0$, 否则 $\delta_i = 1$. 这种情况为响应变量缺失. 类似可以定义协变量缺失. 在实际问题中出现的数据相当一部分是 MAR 的. 所谓 MAR 是指在给定协变量 X 的条件下, Y 与 δ 相互独立, 即

$$P(\delta = 1|Y, X) = P(\delta = 1|X), \tag{1.6.2}$$

(1.6.2) 式表明 Y 是否缺失与 Y 的取值无关, 仅与相应的协变量有关. 随机缺失是统计分析中的一般假定.

缺失数据模型 (1.6.1) 最早是在 Neyman(1938) 的两阶段抽样过程中提出来的, 而后 Rosenbaum 和 Rublin(1943) 与 Cochran(1963) 作了更进一步的补充研究. 在人口统计和经济分析等研究领域中, 用两阶段抽样法可以很好地弥补实际样本缺失的损失，例如在估计 Y 的均值时, 使用该方法可以大大地提高估计的精度.

对于缺失数据, 如果仅用可以完全观测到的数据进行统计推断, 得到的估计量往往会产生偏差, 而且一般不是渐近有效的. 因此, 通常的回归统计推断方法不能直接使用. 解决这个问题的常用方法有两个：一是 Dempster 等 (1977) 提出的 EM 算法. 二是对缺失值进行借补. 这两个方法给缺失数据的统计分析提供了很大帮助, 许多学者在该领域做了大量的研究工作, 并取得了丰硕成果.

1.6.3 删失数据

在生存分析中, 由于实验设计、观测时间等因素的局限, 所要收集的数据往往

不能完全观测. 例如, 在医学实验中, 研究时间往往是固定的, 病人或多或少是以随机的方式进入研究队列中的. 一些人可能在研究结束之前死亡，一些人可能在研究结束之前提前退出或者失去联系, 这便产生了删失数据 (censored data). 概括地说, 当只有部分个体的存活时间能被我们准确知道, 而其余个体的存活时间只知道其发生在某些特定的时间之后就产生删失. 删失通常分为右删失、左删失和区间删失. 一个观测在 L 处右删失指的是其精确值不知道, 但知道其大于或等于 L.

1.6.4 测量误差数据

在许多实际应用中, 往往由于某种原因而使得数据不能精确观测, 而是含有测量误差. 在统计研究中, 通常把带有测量误差的模型称为 EV(errors-in-variables) 模型, 也称为测量误差模型. 在回归分析中, 对响应变量含有测量误差的情况处理比较简单, 可以把测量误差吸收到模型误差中进行处理. 因此, 目前大部分文献集中在协变量带有测量误差的情形, 其研究成果在线性测量误差的情况较多. 所谓线性测量误差是指 $W = X + U$, 其中 X 为不能直接观测的潜在协变量, 而 W 可以直接观测, U 为零均值的测量误差.

参 考 文 献

Allen D M. 1974. The relationship between variable and data augmentation and a method of prediction. Technometrics, 16: 125～127

Antoniadis A. 1997. Wavelets in statistics: A review(with discussion). Journal Italian Statistics Assocasion, 6: 97～144

Antoniadis A, Fan J. 2001. Regularization of wavelets approximations. Journal of the American Statistical Association, 96: 939～967

Baggerly K. 1998. Empirical likelihood as a goodness-of-fit measure. Biometrika, 85(3): 535～547

Cai J W, Fan J, Li R Z, et al.. 2005. Variable selection for multivariate failure time data. Biometrika, 92: 303～316

Chen J, Sitter R R. 1999. A pseudo-empirical likelihood approach to the effective use of auxiliary information in complex surveys. Statistica Sinica, 9: 385～406

Chen S X. 1993. On the accuracy empirical likelihood ratio confidence regions for linear regression model. Annals of The Institute of Statistical Mathematics, 45: 621～637

Chen S X. 1994. Empirical likelihood confidence intervals for linear regression model coefficients. Journal of Multivariate Analysis, 49: 24～40

Chen S X, Cui H J. 2003. An extended empirical likelihood for generalized linear models. Statistica Sinica, 13: 69～81

Chen S X, Hall P. 1993. Smoothed empirical likelihood confidence intervals for quantiles. The Annals of Statistics, 21: 1166～1181

Chen S X, Qin Y S. 2000. Empirical likelihood confidence intervals for local linear smoothers. Biometrika, 47: 946～953

Cochran W. G. 1963. Sampling Techniques (2nd edition). New York: John Wiley

Corcoran S A. 1998. Bartlett adjustment of empirical discrepancy statistics. Biometrika, 85(4): 967～972

Craven P, Wahha G. 1979. Smoothing noisy data with spline functions. Numerische Mathematik, 31: 377～403

Cui H J, Chen S X. 2003. Empirical likelihood confidence region for parameter in the errors-in-variables models. Journal of Multivariate Analysis, 84: 101～115

Cui H J, Kong E F. 2006. Empirical likelihood confidence region for parameters in semi-linear errors-in-variables models. Scandinavian Journal Statistics, 33: 153～168

Cui H J, Li R C. 1998. On parameter estimation for semi-linear errors-in-variables models. Journal of Multivariate Analysis, 64: 1～24

Cui H J, Zhu L X. 2003. Semiparametric regression model with errors in variables. Scandinavian Journal Statistics, 3: 429～442

Dempster A P, Laird N, Rublin D B. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 39: 1～38

Devroye L, Wagner T J. 1980. Distribution-free consistency results in nonparametric discrimination and regression estimation. The Annals of Statistics, 8: 231～239

Diciccio T J, Hall P, Romano J P. 1991. Bartlett adjustment for empirical likelihood. The Annals of Statistics, 19: 1053～1061

Diciccio T J, Hall P, Romano J P. 2000. Comparison of parametric and empirical likelihood functions. Biomentrika, 76(3): 465～476

Diggle P J, Heagerty P, Liang K Y, et al.. 2002. Analysis of Longitudinal Data (2nd edition). Oxford: Oxford University Press

Donoho D L, Johnstone I M. 1995. Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association, 90: 1200～1224

Donoho D L, Johnstone I M, Kerkyacharian G, et al.. 1995. Wavelet shrinkage: asymptopia? Journal of the Royal Statistical Society, Series B, 57: 301～369

Eubank R L. 1988. Spline Smoothing and Nonparametric Regression. New York: Marcel Dekker

Fan J. 1992. Design adaptive nonparametric regression. Journal of the American Statistical Association, 87: 998～1004

Fan J. 1993. Local linear regression smoothers and their minimax efficiency. The Annals of Statistics, 21: 196～216

Fan J, Gijbels I. 1995. Data-driven bandwidth selection in local polynomial fitting: Variable

bandwidth and spatial adaptation. Journal of the Royal Statistical Society, Series B, 57: 371～394

Fan J, Gijbels I. 1996. Local Polynomial Modeling and its Applications. London: Chapman and Hall

Fan J, Heckman N E, Wand M P. 1995. Local polynomial kernel regression for generalized linear models and quasilikelihood functions. Journal of the American Statistical Association, 90: 141～150

Fan J, Li R Z. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96: 1348～1360

Fan J, Li R Z. 2004. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. Journal of the American Statistical Association, 99: 710～723

Fan J, Lv J C. 2009. Sure independence screening for ultra-high dimensional feature space. Journal of the Royal Statistical Society, Series B, 70: 849～911

Fan J, Peng H. 2004. Nonconcave penalized likelihood with a diverging number of parameters. The Annals of Statistics, 32: 928～961

Frank I E, Friedman J H. 1993. A statistical view of some chemometrics regression tools. Technometrics, 35: 109～148

Gasser T, Müller H G. 1979. Kernel estimation of regression function. In Smoothing Techniques for Curve Estimation, Lecture Notes in Mathematics, 757: 23～68. New York: Springer-Verlag

Green P J, Silverman B W. 1994. Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach. London: Chapman and Hall

Hall P, Owen A B. 1993. Empirical likelihood confidence bands in density estimation. Journal of Computing Graphic and Statistics, 2: 273～289

Härdle W. 1990. Applied Nonparametric Regression. Cambridge: Cambridge University Press

Härdle W, Müller M, Sperlich S, et al.. 2004. Nonparametric and Semiparametric Models. Berlin: Springer-Verlag

Huang Z S, Zhang R Q. 2009. Empirical likelihood for nonparametric parts in semiparametric varying-coefficient partially linear models. Statistics & Probability Letters, 79: 1798～1808

Jing B Y, Wood A T A. 1996. Exponential empirical likelihood is not bartlett correctable. The Annals of Statistics, 24(1): 365～369

Kitamura Y. 1997. Empirical likelihood method with weakly dependent processes. The Annals of Statistics, 25: 2084～2102

Koenker R. 2005. Quantile Regression. New York: Cambridge University Press

Kolaczyk E D. 1994. Empirical likelihood for generalized linear models. Statistica Sinica,

4: 199~218

Koopergerg C, Stone C J. 1991. A study of logspline density estimation. Computing Statistics and Data Analysis, 12: 327~347

Koopergerg C, Stone C J, Truong Y K. 1995a. Logspline estimation of a possibly mixed spectral distribution. Journal of Time Series Analysis, 16: 359~388

Koopergerg C, Stone C J, Truong Y K. 1995b. Rate of convergence for logspline spectral density estimation. Journal of Time Series Analysis, 16: 389~401

Li G, Wang Q H. 2003. Empirical likelihood regression analysis with right censored data. Statistica Sinica, 13(1): 51~68

Li G R, Zhu L X, Xue L G, et al.. 2010. Empirical likelihood inference in partially linear single-index models for longitudinal data. Journal of Multivariate Analysis, 101: 718~732

Li R Z, Liang H. 2008. Variable selection in semiparametric regression modeling. The Annals of Statistics, 36: 261~286

Liang H, Li R Z. 2009. Variable selection for partially linear models with measurement errors. Journal of the American Statistical Association, 104: 238~248

Loftsgaarden D O, Quesenberry G P. 1965. A nonparametric estimate of a multivariate density function. Annals of Mathematical Statistics, 36: 1049~1051

Mallows G L. 1973. Some comments on C_p. Technometrics, 15: 661~675

Messer K. 1991. A comparison of a spline estimate to equivalent kernel estimate. The Annals of Statistics, 19: 817~829

Müller H G. 1988. Nonparametric Regression Analysis of Longitudinal Data. Lecture Notes in Statistics, 46. New York: Springer-Verlag

Müller H G. 1991. Smooth optimum kernel estimators near endpoints. Biometrika, 78: 521~531

Mykland P A. 1995. Dual likelihood. The Annals of Statistics, 23: 396~421

Mykland P A. 1999. Bartlett identities and large deviations in likelihood theory. The Annals of Statistics, 27: 1105~1117

Nadaraya E A. 1964. On estimating regression. Theory Probability and Application, 10: 186~190

Neyman J. 1938. Contribution to the theory to sampling human populations. Journal of the American Statistical Association, 33: 101~116

Owen A B. 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika, 74: 237~249

Owen A B. 1990. Empirical likelihood confidence regions. The Annals of Statistics, 18: 90~120

Owen A B. 1991. Empirical likelihood for linear models. The Annals of Statistics, 19: 1725~1747

Owen A B. 2001. Empirical Likelihood. New York: Chapman & Hall

Pagan A, Ullah A. 1999. Nonparametric Econometrics. Cambridge: Cambridge University Press

Pang Z, Xue L G. 2012. Estimation for the single-index models with random effects. Computational Statistics & Data Analysis, 56: 1837～1853

Peng L. 2004. Empirical-likelihood-based confidence interval for the mean with a heavy-tailed distribution. The Annals of Statistics, 32: 1192～1214

Qin J. 1993. Empirical likelihood in based sample problems. The Annals of Statistics, 21: 1182～1196

Qin J, Lawless J. 1994. Empirical likelihood and general estimating equations. The Annals of Statistics, 22: 300～325

Qin J, Zhang B. 2007. Empirical-likelihood-based inference in missing response problems and its application in observational studies. Journal of The Royal Statistical Society, Series B, 69: 101～122

Rosenbaum P R, Rublin D B. 1943. The central role of the propensity score in observational studies for causal effects. Biometrika, 70: 41～55

Ruppert D. 1997. Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. Journal of the American Statistical Association, 92: 1049～1062.

Silverman B W. 1984. Spline smoothing: the equivalent variable kernel method. The Annals of Statistics, 12: 898～916

Shi J, Lau T S. 2000. Empirical likelihood for partially linear models. Journal of Multivariate Analysis, 41: 425～433

Stute W, Xue L G, Zhu L X. 2007. Empirical likelihood inference in nonlinear error in covariables models with validation data. Journal of the American Statistical Association, 102: 332～346

Thall P, Vail S C. 1990. Some covariance models for longitudinal count data with over dispersion. Biometrics, 46: 657～671

Thomas D R, Grunkemeier G L. 1975. Confidence interval estimation of survival probabilities for censored data. Journal of the American Statistical Association, 70: 865～871

Thomas J D, Joseph P R. 1989. On Aajustments based on the signed of the empirical likelihood ratio statistic. Biometrika, 76(3): 447～456

Tibshirani R. 1996. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society, Series B, 58: 267～288

Wahba G. 1977. A survey of some smoothing problems and the method of generalized cross-validation for solving them. In Applications of Statistics (P. R. Krisnaiah, ed.), 507～523. North Holland, Amsterdam

Wahba G. 1990. Spline Models for Observational Data. SIAM, Philadeiphia

Wahba G, Wang Y. 1990. When is the optimal regularization parameter insensitive to the choice of the loss function? Communications of Statistics, 5: 1685～1700

Wang H S, Li R Z, Tsai C L. 2007. Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika, 94: 553～568

Wang L, Li H, Huang J Z. 2008. Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. Journal of the American Statistical Association, 103: 1556～1569

Wang Q H, Linton O, Härdle W. 2004. Semiparametric regression analysis with missing response at random. Journal of the American Statistical Association, 99: 334～345

Wang Q H, Jing B Y. 1999. Empirical likelihood for partial linear models with fixed design. Statistics & Probability Letter, 411: 425～433

Wang Q H, Jing B Y. 2001. Empirical likelihood for a class of functionals of survival distribution with censored data. Annals of The Institute of Statistical Mathematics, 53(3): 517～527

Wang Q H, Rao J N K. 2001. Empirical likelihood for linear regression models under imputation for missing responses. Canadian Journal of Statistics, 29: 597～608

Wang Q H, Rao J N K. 2002a. Empirical likelihood-based inference in linear errors-in-covariables models with validation data. Biometrika, 89: 345～358

Wang Q H, Rao J N K. 2002b. Empirical likelihood-based inference under imputation for missing responses data. The Annals of Statistics, 30: 1～29

Wang Q H, Rao J N K. 2002c. Empirical likelihood-based inference in linear models with missing data. Scandinavian Journal of Statistics, 29: 563～576

Wang Q H, Sun Z H. 2007. Estimation in partially linear models with missing responses at random. Journal of Multivariate Analysis, 98: 1470～1493

Wang Y G, Lin X, Zhu M. 2005. Robust estimating functions and bias correction for longitudinal data analysis. Biometrics, 61: 684～691

Watson G S. 1964. Smooth regression analysis. Sankhya, Series A, 26: 359～372

Xie H L, Huang J. 2009. SCAD-penalized regression in high-dimensional partially linear models. The Annals of Statistics, 37: 673～696

Xue D, Xue L G, Cheng W H. 2011. Empirical likelihood for generalized linear models with missing responses. Journal of Statistical Planning and Inference, 141(6): 2007～2020

Xue L G. 2009a. Empirical likelihood confidence intervals for response mean with data missing at random. Scandinavian Journal of Statistics, 36(4): 671～685

Xue L G. 2009b. Empirical likelihood for linear models with missing responses. Journal of Multivariate Analysis, 100(7): 1353～1366

Xue L G. 2010. Empirical likelihood local polynomial regression analysis of clustered Data. Scandinavian Journal of Statistics, 37(4): 644～663

Xue L G, Xue D. 2011. Empirical likelihood for semiparametric regression model with

missing response data. Journal of Multivariate Analysis, 102(4): 723~740

Xue L G, Zhu L X. 2005. Empirical likelihood confidence regions of the parameters in a partially linear single-index model. Science in China, Series A, 48:1333~1348

Xue L G, Zhu L X. 2006. Empirical likelihood for single-index models. Journal of Multivariate Analysis, 97: 1295~1312

Xue L G, Zhu L X. 2007a. Empirical likelihood for a varying coefficient model with longitudinal data. Journal of the American Statistical Association, 102(478): 642~654

Xue L G, Zhu L X. 2007b. Empirical likelihood semiparametric regression analysis for longitudinal data. Biometrika, 1~16

You J H, Chen G M, Zhou Y. 2006. Block empirical likelihood for longitudinal partially linear regression models. Canadian Journal of Statistics-Revue Canadienne de Statistique, 34(1): 79~96

You J H, Zhou Y. 2006. Empirical likelihood for semiparametric varying-coefficient partially linear regression models. Statistics & Probability Letters, 76: 412~422

Zeger S L, Diggle P J. 1994. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics, 50: 689~699

Zhu L P, Zhu L X. 2009. Nonconcave penalized inverse regression in single-index models with high dimensional predictors. Journal of Multivariate Analysis, 100: 862~875

Zhu L X, Xue L G. 2006. Empirical likelihood confidence regions in a partially linear single-index model. Journal of the Royal Statistical Society, Series B, 68: 549~570

Zou H. 2006. The adaptive LASSO and its oracle properties. Journal of the American Statistical Association, 101: 1418~1429

Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B, 67: 301~320

Zou H, Hastie T, Tibshirani R. 2007. On the degrees of freedom of the LASSO. The Annals of Statistics, 35: 2173~2192

Zou H, Li R Z. 2008. One-step sparse estimates in nonconcave penalized likelihood models. The Annals of Statistics, 36: 1509~1533

Zou H, Yuan M. 2008. Composite quantile regression and the oracle model selection theory. The Annals of Statistics, 36: 1108~1126

Zou H, Zhang H H. 2009. On the adaptive elastic-net with a diverging number of parameters. The Annals of Statistics, 37: 1733~1751

第 2 章　部分线性模型

2.1 引　　言

Engle 等 (1986) 在分析电力需求与气候变化之间的关系时提出了部分线性模型, 该模型具有形式

$$Y_i = \beta_0^{\mathrm{T}} X_i + g(U_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n, \tag{2.1.1}$$

其中 β_0 是 p 维参数向量, $g(\cdot)$ 是未知函数. $\{(X_i, U_i, Y_i), 1 \leqslant i \leqslant n\}$ 是来自 (X, U, Y) 的独立同分布 (iid) 样本, ε_i 是随机误差, 且几乎处处有 $E(\varepsilon_i|X_i, U_i) = 0$. 一般情况下, 设计点列 $\{(X_i, U_i), 1 \leqslant i \leqslant n\}$ 可以是随机的也可以是固定的. U_i 可以是多维变量, 但随着 U_i 的维数增加, 获得 $g(\cdot)$ 的估计所需要的样本量将会非常大. 然而在现实中往往难以提供这种大数据量. 为了避免这一现象发生, 通常假定 U_i 的维数为 1. 不妨设 $U_i \in [0, 1]$, 此时 $g(\cdot)$ 为定义在 $[0, 1]$ 上的未知函数.

模型 (2.1.1) 由两部分构成: 第一部分 $\beta_0^{\mathrm{T}} X_i$ 表示 Y_i 与 X_i 是线性关系; 第二部分 $g(U_i)$ 表明 Y_i 与 U_i 是未知的非线性关系. 如果使用非参数回归来处理, 将会失去 Y_i 与 X_i 线性关系的信息, 从而导致估计值存在较大偏差. 如果用参数回归来拟合, 通常很难得到较好的估计. 因为 Y_i 与 U_i 之间没有明确的函数关系. 此时, 部分线性模型能够有很好的效果. 由于部分线性模型结合了线性模型和非参数模型的特点, 从而具有更大的灵活性. 文献中对这种模型的研究方法一般都是融合了参数回归和非参数回归中常用的方法.

自从部分线性模型问世以来, 已引起了众多学者的关注, 取得了丰富的研究成果. Robinson(1988) 在非参数分量 $g(\cdot)$ 取 Nadaraga-Waston 核估计时, 构造了参数分量 β_0 的加权最小二乘估计 $\hat{\beta}$ 和非参数分量 $g(\cdot)$ 的估计 $\hat{g}(\cdot)$, 在一些正则条件下, 研究了 $\hat{\beta}$ 的渐近正态性和 $\hat{g}(\cdot)$ 的收敛速度.

Speckman(1988) 采用参数化形式 $W\gamma$ 逼近非参数分量 $g(\cdot)$, 其中 W 为某个给定的 $n \times q$ 的满秩矩阵, γ 是附加的 $q \times 1$ 的未知参数向量, 然后利用最小二乘法构造 β_0 的估计. 在适当的条件下, Speckman(1988) 研究了该估计量的渐近性质. 由于在 β_0 估计的构造上已经消除了 U 对 X 和 Y 的影响, 因此 β_0 的估计是渐近无偏的.

一些作者使用光滑样条方法构造了 β_0 和 $g(\cdot)$ 的估计量, 该方法既考虑到估计量同数据的拟合, 又顾及非参数分量估计的光滑性. Heckman(1986) 在 X_i 和 U_i 独

立的情况下研究了 β_0 的惩罚最小二乘估计的相合性和渐近正态性. 其他作者有 Engle 等 (1986), Rice(1986), Green 和 Silverman(1994), Eubank 等 (1998), 他们在不同的条件下分别研究了估计量的大样本性质. Schimek(2000) 对光滑样条方法给出了一个算法. Chen 和 Shiau(1991) 在光滑参数为一个以适当速度趋于 0 的确定序列下, 证明了两阶段样条光滑方法和偏回归方法导致参数分量的有效估计量. 当光滑参数由 GCV 选择时, Chen 和 Shiau(1991) 研究了他们提出的两种有效估计的渐近性质.

Cuzick(1992) 利用半参数似然所得到的渐近估计方程来估计感兴趣的参数, 并证明了所构造的估计量的 $\sqrt{n}$ 相合性. Gao(1995) 证明了参数分量的最小二乘估计的渐近正态性和强相合性. Hamilton 和 Truong(1997) 采用局部线性回归构造了参数和非参数分量的估计, 并证明了估计量的渐近正态性. Bhattacharya 和 Zhao(1997) 在很宽松的假定下构造了参数分量的一个更有效的估计量. Mammen 和 van de Geer(1997) 应用经验过程理论构造了参数的惩罚拟似然估计, 并推导了该估计量的渐近性质. Xue 等 (2004) 提出了参数分量和非参数分量的 sieve 极大似然估计, 证明了参数分量估计的强相合性和渐近正态性, 并得到了非参数函数估计的最优收敛速度. Ma 等 (2006) 研究了异方差的部分线性回归模型, 通过对加权估计方程进行修正, 构造了参数分量和非参数分量的相合估计, 并证明了参数分量的估计是半参数有效的且具有渐近正态性. Liang(2006) 讨论了部分线性模型中的几种估计方法, 并通过数值计算进行了比较. Dabo-Niang 和 Guillas(2010) 研究了具有自回归误差的函数半参数部分线性模型, 构造了参数分量和非参数分量的估计量, 并证明了所提出的估计量的相合性和渐近正态性. Gao 和 Shi(1997) 构造了兴趣参数的估计量, 并证明了它们的相合性和渐近正态性. 其他研究工作还有 (Ai et al., 2011; Shen et al., 2011; You et al., 2011) 等. 对相依数据的研究成果包括 (Gao and Tong, 2004; Aneiros-Pérez and Vilar-Fernández, 2008; Liang and Jing, 2009; Shao, 2009) 等. 用 bootstrap 方法研究部分线性模型的成果有 (Liang et al., 2000; You and Zhou, 2005; Muller and Vial, 2009; Xue and Liu, 2010) 等.

Severini 和 Staniswalis(1994) 与 Härdle 等 (1998) 研究了模型 (2.1.1) 的推广形式: 广义部分线性模型. 为了估计 β_0 和 $g(\cdot)$, Severini 和 Staniswalis(1994) 引进了拟似然估计方法, 该方法有类似于似然函数的性质, 但仅需指定 Y 的二阶矩而不是完全分布. 基于 Severini 和 Staniswalis(1994) 的方法, Härdle 等 (1998) 研究了 $g(\cdot)$ 的线性检验问题. 他们对检验问题的研究补充了 Severini 和 Staniswalis(1994) 的工作. Boente 等 (2006) 针对广义部分线性模型构造了参数和非参数分量的稳健估计量, 并证明了参数分量的估计量的 $\sqrt{n}$ 相合性和渐近正态性.

20 世纪 90 年代以来, 中文期刊发表了许多关于部分线性模型的学术论文. 洪圣岩 (1991) 综合最近邻和最小二乘方法定义了参数和非参数分量的估计量, 证明了参

数分量估计的渐近正态性, 并得到了非参数分量估计的最优收敛速度. Liang(1992) 系统地研究了多种场合下 β_0 的渐近有效估计的构造. Shi(1992) 利用分段多项式逼近方法得到了 β_0 和 $g(\cdot)$ 的稳健 M 估计 $\hat{\beta}$ 和 $\hat{g}(\cdot)$, 在一定条件下证明了 $\hat{\beta}$ 具有渐近正态性, 并得到了 $\hat{\beta}$ 和 $\hat{g}(\cdot)$ 的弱收敛速度. 洪圣岩和赵忠柏 (1993) 综合了核和最小二乘的方法定义了参数和非参数分量的估计量, 证明了参数分量的渐近正态性, 并得到了非参数分量的最优收敛速度. 洪圣岩和成平 (1993) 研究了半参数回归模型的参数估计的 bootstrap 逼近问题. 崔恒建 (1995) 研究了半参数 EV 模型的参数估计理论. 柴根象和徐克军 (1999) 与钱伟民和柴根象 (1999) 将小波方法引入到部分线性模型, 建立了回归参数 β_0 和未知函数 $g(\cdot)$ 的小波估计, 证明了它们具有优良的大样本性质, 并讨论了误差方差的小波估计及其渐近性质. Hong 和 Cheng(1999) 研究了 β_0 的核光滑估计量 $\hat{\beta}$ 正态逼近的精度, 并证明了 $\hat{\beta}$ 的分布不能达到 Berry-Esseen 界的最优速度 $n^{-1/2}$, 除非对 $g(\cdot)$ 的估计量做欠光滑. 薛留根 (2002, 2003) 与 Xue 和 Zhu(2005) 将随机加权方法应用于部分线性模型的研究, 证明了用随机加权统计量的分布逼近原估计量误差的分布的强有效性, 并给出了估计量的最优收敛速度和随机加权逼近速度. 薛留根 (2004, 2005) 研究了部分线性模型中误差方差估计之分布的非一致收敛速度和 Berry-Esseen 界. 薛留根和朱力行 (2007) 研究了纵向数据下部分线性模型的经验似然推断问题. 刘强和薛留根 (2009) 研究了纵向数据下部分线性 EV 模型, 应用一般非参数权函数法和广义最小二乘法给出了参数和非参数分量的估计量, 在一般的条件下, 证明了参数分量估计的渐近正态性, 同时也给出了非参数分量估计的收敛速度, 其结果是独立数据情形下相应结果的推广. 刘强等 (2009) 考虑了在响应变量随机删失情形下的部分线性 EV 模型, 构造了未知参数的两种经验对数似然比函数, 即估计的经验对数似然比和调整的经验对数似然比, 证明了两种经验对数似然比都渐近于 χ^2 分布, 所得结果可以用来构造未知参数的置信域. 其他工作包括 (王启华和郑忠国, 1997; 钱伟民等, 2000; 薛留根和田萍, 2004) 等.

以上成果都是在随机设计点列下进行讨论的. 当 $\{(X_i,U_i),1\leqslant i\leqslant n\}$ 是固定设计时的研究成果相对较少. 而非随机设计情形并不是随机设计情形的特例, 随机设计情况下的结果往往并不能简单地推广到非随机设计情形, 因此二者的处理方法和假设条件也有区别. 这方面的研究成果还有 (Gao and Zhao, 1993; 胡舒合, 1994; 高集体等, 1994, 1995; 王启华, 1995; 石坚, 1998) 等. 他们分别研究了当 $g(\cdot)$ 的估计取一类非参数权函数估计时, β_0 的最小二乘估计和加权最小二乘估计的大样本性质, 其中包括强相合性, 渐近正态性, 收敛速度, Berry-Esseen 界限以及重对数律. 柴根象等 (2001) 在误差 $\{\varepsilon_i,1\leqslant i\leqslant n\}$ 为严平稳 α- 混合时间序列下, 构造了兴趣参数的局部多项式估计, 并研究了估计量的渐近正态性和收敛速度. 柴根象和孙平 (1995) 基于部分线性模型的可加性, 提出了新的二阶段估计方法, 构造了 β_0 和 $g(\cdot)$

的核权函数形式的估计量 $\hat{\beta}$ 和 $\hat{g}(\cdot)$. 在 X_i 为固定设计点列及 U_i 为随机设计点列的情形下, 证明了 $\hat{\beta}$ 的渐近正态性和 $\hat{g}(\cdot)$ 的相合性, 并给出了 $\hat{g}(\cdot)$ 的一致强收敛速度. 关于部分线性模型的详细讨论可进一步参考柴根象和洪圣岩 (1995) 与 Härdle 等 (2000) 的著作.

上述研究成果主要针对完全数据. 由于实际应用的需要, 研究内容不断向各方面发展. 许多作者应用部分线性模型分析了诸如删失数据、测量误差数据、缺失数据、纵向数据等复杂数据, 取得了丰硕的研究成果.

Wang(1996) 和薛留根 (1999) 在核权和近邻权下分别研究了随机删失的部分线性模型, 构造了模型中参数分量和非参数分量的估计量, 并证明了它们的大样本性质. Wang 和 Zheng(1997) 研究了固定设计下随机删失的部分线性模型中兴趣参数的估计及其大样本性质. Chen 和 Khan(2001) 考虑了部分线性删失回归模型, 提出了两阶段估计方法来构造回归参数的估计量, 并证明了所提出的估计量的渐近正态性. 当响应变量是右删失时, Chen 等 (2003) 针对部分线性模型中的回归参数提出了一个简单的估计方法, 证明了所提出的估计量的相合性和渐近正态性.

Liang 等 (1999) 研究了部分线性 EV 模型. 他们使用修正方法使得所得到的参数估计具有相合性和渐近正态性. 崔恒建 (1995) 利用核和广义最小二乘法分别给出 β_0, $g(\cdot)$ 和误差方差的估计量, 获得了估计量的相合性与渐近正态性及其收敛速度. He 和 Liang(2000) 定义了参数分量的分位数回归估计, 并证明了它的渐近正态性. 其他相关工作包括 (Liang, 2000; Ma and Carroll, 2006; Liu, 2007; Liang et al., 2008; You et al., 2008; Wang and Yu, 2007; Zhao and You, 2011) 等.

利用部分线性模型对缺失数据的统计分析也取得了可喜的成果. Liang 等 (2004) 在协变量缺失下对部分线性模型的参数分量和非参数分量提出了新的估计方法, 并讨论了新估计量的渐近有效性. Wang 和 Sun(2007) 考虑了响应变量缺失下的部分线性模型. 他们使用逆概率加权借补方法构造了参数向量和非参数函数的估计量, 证明了参数估计量的渐近正态性, 并给出了非参数估计量的最优收敛速度. Liang 等 (2007) 在响应变量缺失及协变量带测量误差下对兴趣参数提出了一类半参数估计, 并证明了估计量的相合性和渐近正态性. Wang(2009) 考虑了协变量缺失下的部分线性模型, 利用模型校准方法和加权方法来构造参数和非参数分量的估计量, 并证明了估计量的大样本性质. Xue 和 Xue(2011) 利用纠偏技术构造了回归系数的经验似然比统计量, 该统计量渐近于标准 χ^2 分布, 不需要欠光滑回归函数. 其他相关的工作包括 (Liang, 2008, 2009; Sun et al., 2009; Bianco et al., 2010; Liu, 2011; Liu et al., 2011) 等.

近年来, 许多学者研究了纵向数据下部分线性模型. Zeger 和 Diggle(1994) 最早研究了这类模型. 他们提出了后移算法, 得到了参数分量和未知函数的估计量, 并将该方法应用到了 CD4 细胞数据的分析之中. Lin 和 Carroll(2001a) 提出了 profile 核

估计方法, 构造了参数和非参数分量的估计, 并证明了所构造的估计量的大样本性质. Hu 等 (2004) 将 Zeger 和 Diggle(1994) 提出的后移算法与 Lin 和 Carroll(2001a) 提出的 profile 核估计方法进行了比较. 他们通过理论比较发现, 后移算法估计量除了存在偏差问题外, 其渐近方差也比 profile 核方法来得大, 也就是说, 后移算法估计量与 profile 核估计量的渐近效是不同的. Opsomer 和 Ruppert(1997) 对独立数据的研究表明, 两种估计方法具有相同的渐近方差矩阵. 这种现象说明了纵向数据与独立数据的统计分析结果可能存在实质性的差异. He 等 (2002) 将独立数据情形下部分线性模型的 M 估计方法推广到纵向数据情形. 他们在不具体指定误差分布和协方差结构的情况下, 使用 B 样条函数近似非参数函数 $g(\cdot)$, 然后用线性回归模型中的 M 估计方法估计参数向量 β_0 和样条系数向量. 他们在一定条件下证明了 β_0 的估计的相合性、渐近正态性, 并给出了 $g(\cdot)$ 的估计的最优收敛速度. Huang 等 (2007) 研究了纵向数据下广义部分线性模型, 构造了参数分量的有效估计, 并证明了下列性质: 如果正确地指定协方差结构, 所得到的估计量是有效的, 即使错误指定协方差结构, 所得到的估计量仍具有相合性和渐近正态性. 孙孝前和尤进红 (2003) 在固定设计点列下构造了参数分量的迭代加权偏样条最小二乘估计, 并证明了所提出的估计量的渐近正态性, 同时给出了一个自适应方法, 该方法能保证有限次迭代后, 迭代过程会终止, 且产生的估计渐近等价于使用迭代方法所能产生的最好估计. 其他相关文献包括 (Zhang et al., 1998; Lin and Carroll, 2001b; Lin and Ying, 2001; He et al., 2005; Xue and Zhu, 2007; Bai et al., 2008; Qin and Zhu, 2008; Zhou et al., 2008; You and Zhou, 2009; Qin et al., 2009; Mao and Zhu, 2011; Liu, 2011) 等.

一些学者将经验似然方法应用于部分线性模型. Wang 和 Jing(1999) 与 Shi 和 Lau(2000) 研究了固定设计下部分线性模型的经验似然, 证明了所构造的参数的经验对数似然比函数渐近于 χ^2 分布, 由此构造参数的置信域. Wang 和 Jing(2003) 研究了在非固定设计下部分线性模型的经验似然推断问题, 构造了参数 β_0 的经验对数似然比函数, 证明了所构造的经验对数似然比函数渐近于 χ^2 分布. Wang 和 Li(2002) 考虑了右删失下的部分线性模型, 构造了估计的和调整的经验似然统计量并证明了他们的渐近 χ^2 性. Xue 和 Zhu(2007, 2008) 考虑了纵向数据下部分线性模型, 研究了参数向量和基准函数的经验似然问题, 证明了所提出的经验对数似然比函数渐近于标准 χ^2 分布, 由此构造了兴趣参数的置信域和置信区间. Liang 和 Qin(2008) 与 Yang 等 (2009) 分别用经验似然方法研究协变量缺失下的部分线性模型, 构造了参数和非参数分量的置信域和置信区间. Lu(2009) 研究了异方差部分线性模型的经验似然. Liang 等 (2009) 将经验似然方法推广到了广义部分线性模型, 构造了参数和非参数分量的经验似然统计量并证明了它们的渐近 χ^2 性. 其他工作还有 (Qin and Li, 2011a, 2001b; Xue and Xue, 2011; Zhang et al., 2011; Huang,

2012) 等.

上述研究成果主要集中部分线性模型的估计. 关于模型的检验问题的研究较少. Yatchew(1997) 考虑了部分线性模型的非参数部分的检验, 基于差分估计构造了检验统计量. Zhu 和 Ng(2003) 研究了部分线性模型的拟合优度检验问题. 他们构造了基于残差符号经验过程的检验统计量, 并在零假设和对立假设下, 研究了它的性质. Chen 等 (2003) 将 Zhu 和 Ng(2003) 的拟合优度检验问题推广到随机删失数据下的情形, 提出了一个检验统计量来检验数据是否满足一个部分线性模型, 得到了零假设和局部对立假设下检验统计量的渐近分布. You 和 Chen(2005) 研究了部分线性模型误差方差的齐性检验问题, 构造了检验统计量, 并研究了它的渐近性质. Dette 和 Marchlewski(2008) 检验异方差的部分线性模型中误差方差函数是否是参数结构的. Sun 等 (2009) 讨论了响应变量 Y 随机缺失下, 部分线性模型的拟合优度检验. 他们采用借补方法和逆概率加权方法分别构造了基于经验过程的两个检验统计量, 研究了在零假设和局部对立假设下检验统计量的渐近性质. 其他相关工作还有 (Bianco et al., 2006; Pitrun et al., 2006; Li et al., 2011) 等.

目前, 变量选择已经成为统计学研究的热点课题之一, 已广泛应用于金融经济和生物医学等领域. 对参数模型, 已经提出了一系列变量选择方法. 现有的方法大都针对线性回归模型. 对部分线性模型的变量选择, 近几年引起了一些学者的兴趣. Bunea(2004) 利用惩罚最小二乘提出了一个协变量选择方法, 并同时估计参数和非参数分量. Bunea 和 Wegkamp(2004) 提出了两阶段模型选择方法, 其方法实质是首先估计非参数分量, 然后选择参数部分的有效变量. Fan 和 Li(2004) 考虑了纵向数据下部分线性模型的变量选择问题. 他们首先假定参数 β_0 已知, 用局部多项式方法估计非参数分量, 然后采用 SCAD 惩罚技术获得参数分量的 SCAD 惩罚估计, 并证明了惩罚估计具有 oracle 性质. 他们提出的变量选择方法的优点在于变量选择和参数估计可以同时完成. Liang 和 Li(2009) 研究了协变量带有测量误差的部分线性模型的变量选择问题. 他们使用非凸惩罚准则提出了两种变量选择方法: 惩罚最小二乘和惩罚分位数回归. 给出了所得估计量的收敛速度以及 oracle 性质. Xie 和 Huang(2009) 考虑了参数维数发散的部分线性模型的变量选择, 先用多项式样条方法估计非参数函数, 然后采用 SCAD 惩罚得到惩罚估计, 研究了所得估计的收敛速度以及渐近性质. 其他相关的工作有 (Zhu et al., 2003; Avramidis, 2005; Dong et al., 2007; Ni et al., 2009; Cai et al., 2011; Ding et al., 2011; Leng et al., 2011; Fu and Chen, 2012; Ni, 2012) 等.

综上所述, 部分线模型在处理方法上融合了常用的参数方法以及各种非参数方法 (例如, 核估计、局部多项式估计、样条估计等), 但又不是两类方法的简单叠加, 因而其复杂性和难度超过了单一性质的回归模型. 对于部分线模型的研究主要集中在三个方面: 参数和非参数分量的估计, 模型检验, 变量选择. 下面各节从这三个

方面介绍一些代表性成果.

2.2 估 计 方 法

本节结合最小二乘与非参数估计方法来估计部分线性模型中的参数分量和非参数分量. 这里介绍几种主要的估计方法, 其中包括: 权函数方法、偏光滑样条方法、一般光滑方法、分段多项式方法、拟似然方法.

2.2.1 权函数方法

对模型 (2.1.1), 假定 β_0 是已知的, 利用 1.1 节的光滑技术定义 $g(u)$ 的权函数估计

$$\hat{g}(u;\beta_0)=\sum_{i=1}^{n}W_{ni}(u)(Y_i-\beta_0^{\mathrm{T}}X_i), \tag{2.2.1}$$

其中 $\{W_{ni}(u), 1\leqslant i\leqslant n\}$ 为概率权函数, 即满足: $W_{ni}(u)\geqslant 0$, $\sum\limits_{i=1}^{n}W_{ni}(u)=1$. 在 1.2 节中由 (1.2.2) 和 (1.2.10) 式所给出的核权和局部线性权都是很好的选择. 利用 $\hat{g}(\cdot;\beta)$ 可定义 β_0 的最小二乘估计 $\hat{\beta}_{\mathrm{W}}$ 为极小化下列目标函数的解:

$$Q(\beta)=\frac{1}{n}\sum_{i=1}^{n}[Y_i-\beta^{\mathrm{T}}X_i-\hat{g}(U_i;\beta)]^2.$$

简单计算可得

$$\hat{\beta}_{\mathrm{W}}=\left\{\sum_{i=1}^{n}[X_i-\hat{g}_2(U_i)]^{\mathrm{T}}[X_i-\hat{g}_2(U_i)]\right\}^{-1}\sum_{i=1}^{n}[X_i-\hat{g}_2(U_i)][Y_i-\hat{g}_1(U_i)],$$

其中

$$\hat{g}_1(u)=\sum_{i=1}^{n}W_{ni}(u)Y_i,\quad \hat{g}_2(u)=\sum_{i=1}^{n}W_{ni}(u)X_i, \tag{2.2.2}$$

它们分别是 $g_1(u)=E(Y|U=u)$ 和 $g_2(u)=E(X|U=u)$ 的估计量. 利用矩阵形式, 可将 $\hat{\beta}_{\mathrm{W}}$ 表示为

$$\hat{\beta}_{\mathrm{W}}=[(\boldsymbol{X}-\hat{\boldsymbol{g}}_2)^{\mathrm{T}}(\boldsymbol{X}-\hat{\boldsymbol{g}}_2)]^{-1}(\boldsymbol{X}-\hat{\boldsymbol{g}}_2)(\boldsymbol{Y}-\hat{\boldsymbol{g}}_1), \tag{2.2.3}$$

其中 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}$, $\boldsymbol{Y}=(Y_1,\cdots,Y_n)^{\mathrm{T}}$, $\hat{\boldsymbol{g}}_\nu=(\hat{g}_\nu(U_1),\cdots,\hat{g}_\nu(U_n))^{\mathrm{T}}$, $\nu=1,2$. 由 (2.2.1) 式可定义 $g(u)$ 的最终估计量

$$\hat{g}_{\mathrm{W}}(u)\equiv\hat{g}(u;\hat{\beta})=\sum_{i=1}^{n}W_{ni}(u)(Y_i-\hat{\beta}_{\mathrm{W}}^{\mathrm{T}}X_i). \tag{2.2.4}$$

可以用另一种推导方法得到 β_0 和 $g(\cdot)$ 的估计量. 对 (2.1.1) 两边求关于 U_i 的条件期望可得到

$$g(U_i) = g_1(U_i) - \beta_0^{\mathrm{T}} g_2(U_i).$$

把上式代入 (2.1.1) 式, 有

$$Y_i - g_1(U_i) = \beta_0^{\mathrm{T}}[X_i - g_2(U_i)] + \varepsilon_i, \quad i = 1, \cdots, n. \tag{2.2.5}$$

上式意味着将 X_i 和 Y_i 条件中心化, 这样可以消去模型 (2.1.1) 中的截距项. 若将 $g_1(\cdot)$ 和 $g_2(\cdot)$ 看做已知, 那么 (2.2.5) 是一个线性模型. 利用最小二乘法可求得

$$\tilde{\beta} = [(\boldsymbol{X} - \boldsymbol{g}_2)^{\mathrm{T}}(\boldsymbol{X} - \boldsymbol{g}_2)]^{-1}(\boldsymbol{X} - \boldsymbol{g}_2)(\boldsymbol{Y} - \boldsymbol{g}_1), \tag{2.2.6}$$

其中 $\boldsymbol{g}_\nu = (g_\nu(U_1), \cdots, g_\nu(U_n))^{\mathrm{T}}$, $\nu = 1, 2$. 将 (2.2.2) 式所定义的估计量 $\hat{g}_1(\cdot)$ 和 $\hat{g}(\cdot)$ 代入 (2.2.6) 式即可得到 (2.2.3) 式给出的估计量 $\hat{\beta}_{\mathrm{W}}$. 然后利用估计量 $\hat{\beta}_{\mathrm{W}}$, $\hat{g}_1(u)$ 和 $\hat{g}_2(u)$ 可以得到 $g(u)$ 的估计量

$$\hat{g}_{\mathrm{W}}(u) = \hat{g}_1(u) - \hat{\beta}_{\mathrm{W}}^{\mathrm{T}} \hat{g}_2(u).$$

该估计量与 (2.2.4) 式给出的估计量是一样的.

2.2.2 偏光滑样条方法

光滑样条方法可以用来对半参数回归模型中的参数分量和非参数分量进行估计. 关于样条估计较早的文献有 (Engle et al., 1986; Heckman, 1986; Rice, 1986; Eubank et al., 1998). 事实上, 光滑样条估计继承了核估计的某些性质, 但又有差别. 下面给出光滑样条方法的具体描述. 设 $g(\cdot)$ 在 $L^2[0,1]$ 上具有 k 阶连续导数, 那么定义 β_0 和 $g(\cdot)$ 的光滑样条估计为极小化

$$\frac{1}{n}\sum_{i=1}^{n}[Y_i - \beta^{\mathrm{T}} X_i - g(U_i)]^2 + \lambda \int_0^1 [g^{(k)}(u)]^2 \mathrm{d}u \tag{2.2.7}$$

的解, 其中 $\lambda > 0$ 是光滑参数, 它起到在拟合程度与光滑程度之间的平衡作用. 当 $k = 2$ 时, 称上式为立方光滑样条. 由于 (2.2.7) 式是把非参数光滑样条推广到了部分线性模型, 因此人们称这种方法为 "偏光滑样条" 方法. Heckman(1986) 对光滑样条估计的性质进行了研究. 2.2.3 节再对偏光滑样条进行较详细的讨论.

2.2.3 一般光滑方法

1. 偏样条

首先介绍将非参数分量 $g(\cdot)$ 参数化的方法, 然后引入偏样条方法. 对函数空间 $\mathcal{G}$ 施加一定光滑性的限制, 由于 $\mathcal{G}$ 是无穷维的, 通常由光滑性使得 $\mathcal{G}$ 中的元参数

化. 例如 $\mathcal{G}$ 中的元在选定的基 $\{b_j(\cdot), 1 \leqslant j \leqslant q\}$ 下有线性表示, 即

$$g(u) = \sum_{j=1}^{q} b_j(u)\gamma_{j0},$$

其中 q 是待选定的光滑参数. 在选定基 $\{b_j(\cdot), 1 \leqslant j \leqslant q\}$ 和 q 下, 将估计非参数分量 g 的问题转化为参数估计问题. 假设模型 (2.1.1) 中的 g 可参数化为 $\boldsymbol{g} = (g(U_1), \cdots, g(U_n))^{\mathrm{T}} = B\gamma_0$, 其中 B 为 $n \times q$ 的某个给定的矩阵, $\gamma_0 = (\gamma_{10}, \cdots, \gamma_{q0})^{\mathrm{T}}$ 是 $q \times 1$ 参数向量. 为假定 $n \times (p+q)$ 矩阵 $(\boldsymbol{X}, B)$ 非奇异, 我们设向量 $(1, \cdots, 1)^{\mathrm{T}}$ 与 B 共面, 但不与 $\boldsymbol{X}$ 共面. 利用矩阵记号, 可将部分线性模型 (2.1.1) 改写为

$$\boldsymbol{Y} = \boldsymbol{X}\beta_0 + B\gamma_0 + \varepsilon,$$

通过解极小化问题

$$\min_{\beta \in \mathbf{R}^p, \gamma \in \mathbf{R}^q} ||\boldsymbol{Y} - \boldsymbol{X}\beta - B\gamma||^2,$$

可得正规方程组

$$\begin{cases} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\beta = \boldsymbol{X}^{\mathrm{T}}(\boldsymbol{Y} - B\gamma), \\ B\gamma = P_B(\boldsymbol{Y} - \boldsymbol{X}\beta), \end{cases} \tag{2.2.8}$$

其中 $P_B = B(B^{\mathrm{T}}B)^{-1}B$ 为投影矩阵. Green(1985) 提出用一个光滑矩阵 S(也许非线性) 代替方程 (2.2.8) 中的 P_B, 并解方程组

$$\begin{cases} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\beta = \boldsymbol{X}^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{g}), \\ \boldsymbol{g} = S(\boldsymbol{Y} - \boldsymbol{X}\beta), \end{cases}$$

可得 β_0 和 $\boldsymbol{g}$ 的估计量为

$$\hat{\beta}_{\mathrm{PS}} = [\boldsymbol{X}^{\mathrm{T}}(I - S)\boldsymbol{X}]^{-1}\boldsymbol{X}^{\mathrm{T}}(I - S)\boldsymbol{Y}, \tag{2.2.9}$$

$$\hat{\boldsymbol{g}}_{\mathrm{PS}} = S(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{\mathrm{PS}}). \tag{2.2.10}$$

如果 S 是样条光滑矩阵, 那么 (2.2.9) 恰好是极小化 (2.2.7) 式得到的 β_0 的估计量. 因此, 称估计量 $\hat{\beta}_{\mathrm{PS}}$ 和 $\hat{\boldsymbol{g}}_{\mathrm{PS}}$ 为偏样条估计, 并称该估计方法为偏样条方法. 若取 S 为核权函数, 则称所得的估计为 Green-Jennison-Seheult 估计, 也称之为偏核光滑估计. Speckman(1988) 对偏核光滑估计及其大样本性质进行了详细讨论.

2. 偏回归

受到前面介绍的几种方法的启发, 我们给出另一种非参数回归方法 —— 偏回归. 设 $\widetilde{\boldsymbol{X}} = (I - S)\boldsymbol{X}$ 和 $\widetilde{\boldsymbol{Y}} = (I - S)\boldsymbol{Y}$ 为对 X 和 Y 调整后的变量, 且它们依赖

于变量 U, 其中 S 为任意光滑矩阵. 把 $\widetilde{\boldsymbol{Y}}$ 看作 $\widetilde{\boldsymbol{X}}\beta_0$ 的线性回归, 并利用通常的最小二乘法可定义 β_0 的偏回归估计

$$\hat{\beta}_{\mathrm{PR}} = (\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}})^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\boldsymbol{Y}}. \tag{2.2.11}$$

如果 S 为对称幂等矩阵, 那么公式 (2.2.9) 和 (2.2.11) 是一样的, 但在一般情况下二者是有区别的. (2.2.11) 式中 β_0 的估计量实际上是加权最小二乘准则

$$\min_{\beta\in R^p} ||(I-S)(\boldsymbol{Y}-\boldsymbol{X}\beta)||^2$$

的解. 利用估计量 $\hat{\beta}_{\mathrm{PR}}$ 并类似于 (2.2.10) 式可定义 g 的偏回归估计量, 这只需将 (2.2.10) 式中的 $\hat{\beta}_{\mathrm{PS}}$ 换为 $\hat{\beta}_{\mathrm{PR}}$ 即可.

Speckman(1988) 证明了上述两个方法对 g 的拟合达到了类似的效果, 但 Rice (1986) 指出了 (2.2.9) 式的 $\hat{\beta}_{\mathrm{PS}}$ 有偏差问题. 然而, 由 (2.2.11) 式给出的 $\hat{\beta}_{\mathrm{PR}}$ 的偏差是低阶的, 且不需要欠光滑即可渐近可忽略. 如果主要兴趣是对 β_0 作推断, Speckman(1988) 建议使用 $\hat{\beta}_{\mathrm{PR}}$ 更合适.

权函数法与偏回归有密切的联系. 给定一个权函数 $\{W_{ni}(u), 1\leqslant i\leqslant n\}$, 相应的光滑矩阵为

$$S=\begin{pmatrix} W_{n1}(U_1) & W_{n2}(U_1) & \cdots & W_{nn}(U_1)\\ W_{n1}(U_2) & W_{n2}(U_2) & \cdots & W_{nn}(U_2)\\ \vdots & \vdots & & \vdots\\ W_{n1}(U_n) & W_{n2}(U_n) & \cdots & W_{nn}(U_n)\end{pmatrix}.$$

将上述光滑矩阵 S 代入 (2.2.11) 式即可得到 (2.2.3) 式给出的估计量 $\hat{\beta}_{\mathrm{W}}$. 因此, 加权估计量 $\hat{\beta}_{\mathrm{W}}$ 也是偏回归估计量.

2.2.4 分段多项式方法

Chen(1988) 采用了分段多项式最小二乘估计得到参数 β_0 的估计. 不失一般性, 假定 U_i 在区间 $[0,1]$ 上取值. 分段多项式估计的思想如下: 给定与 n 有关的正整数 M_n, 将区间 $[0,1]$ 等分 M_n 份, 并记为 $I_{n\nu}=[(\nu-1)/M_n,\nu/M_n)$, $1\leqslant\nu<M_n$, $I_{nM_n}=[1-1/M_n,1]$. 用 $\psi_{n\nu}(u)$ 表示 $I_{n\nu}$ 的示性函数, 记

$$\varphi_{n\nu}(\mu)=\psi_{n\nu}(u)(1,u-u_\nu,\cdots,(u-u_\nu)^k)^{\mathrm{T}},\quad \nu=1,\cdots,M_n,$$

$$\varphi_n(u)=(\varphi_{n1}^{\mathrm{T}}(u),\cdots,\varphi_{nM_n}^{\mathrm{T}}(u))^{\mathrm{T}},$$

其中 u_ν 可取为 $I_{n\nu}$ 的中点, 也可取为 $I_{n\nu}$ 中别的点. 因此, 可以用区间 $[0,1]$ 上的 k 次分段多项式函数 $\gamma^{\mathrm{T}}\varphi_n(u)$ 来逼近 $g(u)$, 即

$$g(u) \approx \gamma^{\mathrm{T}}\varphi_n(u),$$

其中 γ 是 $q\times 1$ 的参数向量, $q=(k+1)M_n$. β_0 和 γ 的估计定义为极小化目标函数

$$Q_n(\beta,\gamma) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - \beta^{\mathrm{T}}X_i - \gamma^{\mathrm{T}}\varphi_n(U_i)]^2$$

的解. 记所得的解为 $\hat{\beta}_{\mathrm{PP}}$ 和 $\hat{\gamma}$, 则 β_0 和 $g(u)$ 的估计分别为 $\hat{\beta}_{\mathrm{PP}}$ 和 $\hat{g}_{\mathrm{PP}}(u) = \hat{\gamma}^{\mathrm{T}}\varphi_n(u)$. 注意到估计过程中涉及 k, M_n 以及节点 $u_1,\cdots,u_{M_n}$ 的选取问题. 对估计量影响较大的是 M_n, 因此, 需根据具体情况对 M_n 给以适当限制. Chen(1988) 证明了分段多项式所得到的估计 $\hat{\beta}_{\mathrm{PP}}$ 具有 $\sqrt{n}$ 收敛速度.

由于最小二乘法得到的估计量缺乏稳健性, 故 Huber(1981) 提出了稳健估计思想. Shi(1992) 与 Xue 和 Zhu(2005) 利用分段多项式逼近方法得到了 β_0 和 $g(\cdot)$ 的稳健 M 估计, 在一定条件下证明了估计的渐近性质.

2.2.5 拟似然方法

在实际问题中, 响应变量 Y 和协变量 (X,U) 之间的相关性具有形式

$$E(Y|X,U) = H(\beta_0^{\mathrm{T}}X + g(U)), \tag{2.2.12}$$

其中 $X\in\mathbf{R}^p$, $U\in\mathcal{U}\subseteq\mathbf{R}$, β_0 为未知的 p 维参数向量, $g(\cdot)$ 为 $\mathcal{U}$ 上未知的光滑函数, $H(\cdot)$ 为已知函数, 称这个模型为广义部分线性模型. Severini 和 Staniswalis(1994) 与 Härdle 等 (1998) 分别研究了模型 (2.2.12) 的估计和检验问题, Boente 等 (2006) 构造了参数和非参数分量的稳健估计量. 本节主要介绍 Severini 和 Staniswalis(1994) 的拟似然估计方法.

假设 $\mu(X,U) = H(\beta^{\mathrm{T}}X + g(U))$, $\mathrm{var}(Y|X,U) = \sigma^2V(\mu)$, 其中 σ^2 为未知参数, $V(\cdot)$ 为已知函数. 为估计 β_0 和 $g(\cdot)$, 遵循 Severini 和 Staniswalis(1994) 的思想, 取拟似然函数

$$Q(\mu;y) = \int_y^{\mu}\frac{y-t}{V(t)}\mathrm{d}t.$$

设 $\{(X_i,U_i,Y_i), 1\leqslant i\leqslant n\}$ 为来自 (X,U,Y) 的 iid 样本. 对固定的 β_0 和 u, 用 $\hat{g}(u;\beta_0)$ 表示下列关于 η 的加权似然方程的解:

$$\frac{\mathrm{d}}{\mathrm{d}\eta}\sum_{i=1}^{n}Q(H(\beta_0^{\mathrm{T}}X_i+\eta);Y_i)K_h(U_i-u) = 0, \tag{2.2.13}$$

其中 $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ 为核函数, $h = h_n$ 为带宽且 $0 < h \to 0$. 给定估计量 $\hat{g}(u;\beta)$, 定义 β_0 的估计量 $\hat{\beta}$ 为下列方程的解:

$$\frac{\partial}{\partial\beta}\sum_{i=1}^{n} Q(H(\beta^{\mathrm{T}}X_i + \hat{g}(U_i;\beta)); Y_i) = 0. \tag{2.2.14}$$

虽然 (2.2.13) 式可以用来得到 $\hat{g}(u;\beta_0)$, 但所得到的估计量对 u 取接近于 $\mathcal{U}$ 的边界点比远离边界点的偏差更高. 这在实际和理论上都是不便的. 有几个方法可以削减接近于边界的偏差, 一是边界纠偏核估计 (Rice, 1984), 二是局部线性估计 (Fan and Gijbels, 1996), 三是对 (2.2.14) 中组成和的项进行 "修整"(trimming), 仅使用远离边界的观察 U_i. 虽然第三种方法按实用的观点是不可取的, 但其优越性是理论结果可以用清晰的形式表示. 因此, 这里我们使用第三种方法描述估计程序和表示渐近结果.

用 $\mathcal{U}_0 \subseteq \mathcal{U}$ 表示 U 的支撑集. 定义示性 I_i: 如果 $U_i \in \mathcal{U}_0$, 则 $I_i = 1$, 否则 $I_i = 0$. 那么, 求解方程 (2.2.14) 的问题可改为求解

$$\frac{\partial}{\partial\beta}\sum_{i=1}^{n} I_i Q(H(\beta^{\mathrm{T}}X_i + \hat{g}(U_i;\beta)); Y_i) = 0. \tag{2.2.15}$$

在某些情况下, 可以求解 (2.2.13) 或 (2.2.15) 而得到估计量的明确表达式. 例如, 设 $H(t) = t$ 且 $V(\mu) = 1$; 那么, 所得到的拟似然函数是正态分布数据的对数似然函数. 在这种情况下, 求解方程 (2.2.13) 所得到的估计量正是 (2.2.1) 式给出的估计量, 其中权函数是由 (1.2.2) 式定义的核权函数. 同时也可以求解方程 (2.2.14) 而得到 β_0 的估计量

$$\hat{\beta} = [(\boldsymbol{X} - \hat{\boldsymbol{g}}_2)^{\mathrm{T}} D(\boldsymbol{X} - \hat{\boldsymbol{g}}_2)]^{-1}(\boldsymbol{X} - \hat{\boldsymbol{g}}_2)D(\boldsymbol{Y} - \hat{\boldsymbol{g}}_1), \tag{2.2.16}$$

其中 $\hat{\boldsymbol{g}}_1$ 和 $\hat{\boldsymbol{g}}_2$ 在 (2.2.2) 式中定义, D 是 $n \times n$ 对角矩阵, 其第 i 个对角元素为 I_i. 如果 $\mathcal{U}_0 = \mathcal{U} = [0,1]$, 那么 D 为单位矩阵, 此时, (2.2.16) 式就是形如 (2.2.3) 式给出的估计量. 估计量 $\hat{\beta}$ 本质上就是由 Robinson(1988) 和 Speckman(1988) 所研究的估计量.

为估计 σ^2, 我们使用类似于标准的参数拟似然方法. 设 $\hat{H}_i = H(\hat{\beta}^{\mathrm{T}}X_i + \hat{g}(U_i;\hat{\beta}))$, 那么, σ^2 的估计量定义为

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{H}_i)^2/V(\hat{H}_i).$$

Severini 和 Staniswalis(1994) 研究了估计量 $\hat{\beta}$, $\hat{g}$ 和 $\hat{\sigma}^2$ 的渐近正态性, 即在一些正则条件下, 有 $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$,

$$\sup_{u\in\mathcal{U}_0} |\hat{g}(u;\hat{\beta}) - g(u)| = O_P(n^{-1/4}),$$

$$\sqrt{n}(\hat{\beta}-\beta_0) \xrightarrow{D} N(0,\sigma^2\Sigma^{-1}),$$

其中 Σ 为 $p\times p$ 矩阵, 其 (i,j) 元素为

$$E\left[\frac{\partial^2}{\partial\beta_i\partial\beta_j}Q(H(\beta^{\mathrm{T}}X_1+g(U_1;\beta));Y_1)I_1\Big|_{\beta=\beta_0}\right],$$

$g(u;\beta_0)$ 为下列关于 η 的方程的解:

$$E\left[\frac{\mathrm{d}}{\mathrm{d}\eta}Q(H(\beta_0^{\mathrm{T}}X+\eta);Y)\Big|U=u\right]=0.$$

Severini 和 Staniswalis(1994) 也构造了 $\hat{\beta}$ 的渐近方差矩阵的估计. 同时也将提出的方法推广到了 Y 为多维响应变量的情况.

2.2.6 经验似然方法

如果 $g_1(u)$ 和 $g_2(u)$ 是已知的, 那么 (2.2.5) 是关于 β_0 的线性模型, 按照 Owen(1991) 提出的方法可以构造 β_0 的经验对数似然比函数, 其渐近分布为 χ^2 分布. 然而, $g_1(u)$ 和 $g_2(u)$ 通常是未知的, 所构造的经验似然比不能直接用来对 β_0 作统计推断, 因为它包含两个未知函数 $g_1(u)$ 和 $g_2(u)$. 因此, 我们用其估计量代替他们, 并定义 β_0 的经验似然比统计量. 这里使用核方法来估计 $g_1(u)$ 和 $g_2(u)$, 其估计量如同 (2.2.2), 其中取核权函数. 基于上述讨论, 引入辅助随机变量

$$\hat{Z}_i(\beta)=[X_i-\hat{g}_2(U_i)]\{Y_i-\hat{g}_1(U_i)-\beta^{\mathrm{T}}[X_i-\hat{g}_2(U_i)]\},$$

其中 β 为任意参数向量. 那么 β 的经验对数似然比函数定义为

$$\hat{l}(\beta)=-2\max\left\{\sum_{i=1}^{n}\log(np_i)\Big|p_i\geqslant 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\hat{Z}_i(\beta)=0\right\}.$$

Wang 和 Jing(2003) 在适当的条件下证明了 $\hat{l}(\beta_0)$ 以分布收敛于自由度为 p 的 χ^2 变量. 这个结果不但可以对假设 $H_0:\beta=\beta_0$ 进行检验, 也可以用来构造 β_0 的置信域. 其置信域为

$$\{\beta\,|\,\hat{l}(\beta)\leqslant\chi_p^2(1-\alpha)\}.$$

此外, 极大化 $\{-\hat{l}(\beta)\}$ 可以得到 β_0 的极大经验似然估计量 $\hat{\beta}_{\mathrm{MEL}}$. 可以证明 $\hat{\beta}_{\mathrm{MEL}}$ 具有形式 (2.2.3). Wang 与 Jing(1999) 和 Shi 与 Lau(2000) 对固定设计下部分线性模型讨论了上述类似的问题.

2.2.7 模拟

例2.2.1 考虑部分线性模型

$$Y_i = \beta^{\mathrm{T}} X_i + 50(U_i - 0.5)^2 + \varepsilon_i,$$

其中 $\beta = (\beta_1, \beta_2)^{\mathrm{T}} = (0.8, 0.6)^{\mathrm{T}}$, $\{X_i, 1 \leqslant i \leqslant n\}$ 为来自二维标准正态分布的独立随机样本, $\{\varepsilon_i, 1 \leqslant i \leqslant n\}$ 为服从正态分布 $N(0, 0.6^2)$ 的独立随机变量, $g(u) = 50(u - 0.5)^2$.

对 (β_1, β_2) 的估计, 我们使用两个方法做模拟: 2.2.2 小节提出的偏样条 (PS) 和偏回归 (PR) 方法. 对偏样条估计量, 使用三次样条基, 用交叉验证方法选取节点数. 偏回归估计量中的光滑矩阵由局部多项式权构造, 其中核函数取为 $K(u) = 0.75(1-u^2)_+$, 用交叉验证方法选取带宽. 分别在样本量 $n = 50, 80, 150$ 下通过 5000 次运行计算偏差 (Bias)、标准差 (SD) 和均方误差 (MSE). 模拟结果放在表 2.2.1 中.

表 2.2.1 例 2.2.1 的模拟结果. 估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的偏差、标准差和均方误差

n	方法	$\hat{\beta}_1$			$\hat{\beta}_2$		
		Bias	SD	MSE	Bias	SD	MSE
50	PS	0.00167	0.09747	0.00950	0.00008	0.09910	0.00982
	PR	−0.00172	0.12227	0.01495	0.00058	0.12209	0.01491
80	PS	−0.00127	0.07309	0.00534	0.00004	0.07101	0.00504
	PR	−0.00135	0.08370	0.00701	−0.00016	0.08145	0.00663
150	PS	0.00060	0.05121	0.00262	−0.00003	0.05094	0.00260
	PR	0.00071	0.05497	0.00302	−0.00034	0.05448	0.00297

从表 2.2.1 可以看出, β_1 和 β_2 的偏样条估计比偏回归估计有更小的偏差、标准差和均方误差. 这就表明偏样条比偏回归执行效果更好. 此外, 偏差、标准差和均方误差都随着样本量的增加而减少.

2.3 删失数据分析

本节主要介绍删失数据下部分线性模型中参数分量和非参数分量的统计推断, 主要内容包括: 估计方法和经验似然.

2.3.1 估计方法

在可靠性寿命和医药追踪试验等实际问题中, 模型 (2.1.1) 中响应变量 $\{Y_i, 1 \leqslant i \leqslant n\}$ 常常因随机右删失而不能被完全被观察, 我们仅能观察到 $\{(Z_i, \Delta_i), 1 \leqslant i \leqslant n\}$, 其中

$$Z_i = \min\{Y_i, \Delta_i\}, \quad \Delta_i = I(Y_i \leqslant C_i), \quad i = 1, \cdots, n, \tag{2.3.1}$$

这里 $\{C_i, 1 \leqslant i \leqslant n\}$ 为独立于 $\{Y_i, 1 \leqslant i \leqslant n\}$ 的 iid 随机变量序列, 且具有共同的连续分布函数 $G(\cdot)$; 称 C_i 为删失变量. 假设 $\{Y_i, 1 \leqslant i \leqslant n\}$ 具有共同的连续分布函数 $F(\cdot)$, 因此 $\{Z_i, 1 \leqslant i \leqslant n\}$ 具有共同的分布函数 $H = 1-(1-F)(1-G)$. 在响应变量具有右删失情况下, 称模型 (2.1.1) 和 (2.3.1) 为随机右删失的部分线性模型.

我们的问题是如何利用样本 $\{(X_i, U_i, Z_i, \Delta_i), 1 \leqslant i \leqslant n\}$ 估计参数向量 β_0 和未知回归函数 $g(\cdot)$. 为方便起见, 下文约定对任何分布 $V(u)$, 定义 $\tau_V = \inf\{u : V(u) = 1\}$, $\bar{V}(u) = 1 - V(u)$, 对任何 $r > 0$, 定义 $V^{-r}(u) = [V(u)]^{-r}$, 且约定书中出现的 c 表示任何所需要的常数.

用 $\widetilde{F}$ 表示给定 (X_i, U_i) 下 Y_i 的条件分布. 记 $Z_{iG} = Z_i\Delta_i/[1 - G(Z_i)]$. 注意到

$$\begin{aligned}E(Z_{iG}|X_i, U_i) &= \int_{-\infty}^{\tau_{\widetilde{F}}}\int_y^{\tau_G} \frac{y}{1-G(y)}\mathrm{d}G(u)\mathrm{d}\widetilde{F}(y)\\ &= E(Y_i|X_i, U_i) = \beta_0^{\mathrm{T}}X_i + g(U_i), \quad i = 1, \cdots, n.\end{aligned}$$

由此可以认为 $\{Z_{iG}, 1 \leqslant i \leqslant n\}$ 服从模型

$$Z_{iG} = \beta_0^{\mathrm{T}}X_i + g(U_i) + e_i, \quad i = 1, \cdots, n, \tag{2.3.2}$$

其中 $\{e_i, 1 \leqslant i \leqslant n\}$ 相互独立且均值为 0. 当 $G(\cdot)$ 已知时, 利用样本 $\{(X_i, U_i, Z_{iG}), 1 \leqslant i \leqslant n\}$ 并按照 2.2 节的方法可以构造 β_0 和 $g(u)$ 的估计量. 然而, $G(\cdot)$ 通常是未知的, 我们需要估计它. 以 $1 - G(\cdot)$ 的 Kaplan-Meier 估计

$$1 - \hat{G}(z) = \prod_{i=1}^{n}\left(\frac{N^+(Z_i)}{1 + N^+(Z_i)}\right)^{I(Z_i \leqslant z, \Delta_i = 0)}, \quad -\infty < z < \infty \tag{2.3.3}$$

作为 $1 - G(z)$ 的估计, 其中

$$N^+(z) = \sum_{j=1}^{n} I(Z_j > z).$$

使用 2.2.1 小节给出的权函数法, 并用 $Z_{i\hat{G}}$ 代替 Y_i 可以得到 β_0 和 $g(u)$ 的估计量 $\hat{\beta}$ 和 $\hat{g}(u)$, 其计算公式分别为

$$\hat{\beta}_{\mathrm{C}} = [(\boldsymbol{X} - \hat{\boldsymbol{g}}_2)^{\mathrm{T}}(\boldsymbol{X} - \hat{\boldsymbol{g}}_2)]^{-1}(\boldsymbol{X} - \hat{\boldsymbol{g}}_2)(\boldsymbol{Z}_{\hat{G}} - \hat{\boldsymbol{g}}_{1\hat{G}}) \tag{2.3.4}$$

和 $\hat{g}_{\mathrm{C}}(u) = \hat{g}_{1\hat{G}}(u) - \hat{\beta}_{\mathrm{C}}^{\mathrm{T}}\hat{g}_2(u)$, 其中 $\boldsymbol{X} = (X_1, \cdots, X_n)^{\mathrm{T}}$, $\boldsymbol{Z}_{\hat{G}} = (Z_{1\hat{G}}, \cdots, Z_{n\hat{G}})^{\mathrm{T}}$, $\hat{\boldsymbol{g}}_{1\hat{G}} = (\hat{g}_{1\hat{G}}(U_1), \cdots, \hat{g}_{1\hat{G}}(U_n))^{\mathrm{T}}$, $\hat{\boldsymbol{g}}_2 = (\hat{g}_2(U_1), \cdots, \hat{g}_2(U_n))^{\mathrm{T}}$,

$$\hat{g}_{1\hat{G}}(u) = \sum_{i=1}^{n} W_{ni}(u)Z_{i\hat{G}}, \quad \hat{g}_2(u) = \sum_{i=1}^{n} W_{ni}(u)X_i. \tag{2.3.5}$$

这里 $\hat{g}_{1\hat{G}}(u)$ 和 $\hat{g}_2(u)$ 分别为 $g_1(u) = E(Y|U = u)$ 和 $g_2(u) = E(X|U = u)$ 的估计量, $\{W_{ni}(u), 1 \leqslant i \leqslant n\}$ 为某个概率权函数. 当 X_i 的维数 $p = 1$ 时, Wang 和 Zheng(1997) 在核权的情况下证明了 $\hat{\beta}_C$ 的渐近正态性, 并给出了 $\hat{g}_C(u)$ 的最优收敛速度为 $O_P(n^{-1/3})$; 薛留根 (1999) 在近邻权的情况下证明了 $\hat{\beta}_C$ 的渐近正态性, 同时得到了 $\hat{g}_C(u)$ 的最优收敛速度也为 $O_P(n^{-1/3})$. 在任意维数 p 下, 两种权函数所得到的 $\hat{\beta}_C$ 的渐近方差皆为 $n^{-1}\Gamma_1^{-2}\Gamma_2$, 其中

$$
\begin{aligned}
\Gamma_1 &= E\{[X_1 - E(X_1|U_1)]^{\otimes 2}\}, \\
\Gamma_2 &= E\{[X_1 - E(X_1|U_1)]^{\otimes 2}[Z_{1G} - \beta_0^{\mathrm{T}} X_1 - g(U_1)]^2\} \\
&\quad + 2E\{[X_1 - E(X_1|U_1)][X_2 - E(X_2|U_2)]^{\mathrm{T}}\eta_0(Z_1, \Delta_1; Z_2) Z_{1G} Z_{2G}\} \\
&\quad + E\{E[(X_2 - E(X_2|U_2))\eta_0(Z_1, \Delta_1; Z_2) Z_{2G}|Z_1, \Delta_1]\}^{\otimes 2}, \\
\eta_0(Z_1, \Delta_1; z) &= \int_0^{Z_1 \wedge z} \bar{H}^{-2}(u)\mathrm{d}\widetilde{H}_0(u) + \bar{H}^{-1}(Z_1) I(Z_1 \leqslant z, \Delta_1 = 0).
\end{aligned}
$$

这里 $H(z) = P(Z_1 \leqslant z)$, $\widetilde{H}_0(u) = P(Z_1 > 0, \Delta_1 = 0)$, $Z_1 \wedge z = \min\{Z_1, z\}$, $A^{\otimes 2}$ 表示 AA^{T}.

如果没有右删失 (即完全数据), 则可看作 $C_1 = \infty$. 此时, $G(\cdot) = 0$, $\hat{G}(\cdot) = 0$, $\eta_0(Z_1, \Delta_1; Z_2) = 0$. 从而, $Z_{i\hat{G}} = Y_i$, $n^{-1}\Gamma_1^{-2}\Gamma_2 = n^{-1}\sigma^2\Gamma_1^{-1}$, 它即为完全数据情形下的渐近方差. 由此可见, 渐近方差 $n^{-1}\Gamma_1^{-2}\Gamma_2$ 正是完全数据下相应结果的推广.

2.3.2 经验似然

1. 估计的经验似然

由模型 (2.3.2), 并利用 2.2.6 小节的方法, 可以构造任何参数 β 的经验似然比函数. 为此, 引入辅助随机变量

$$
\hat{\eta}_i(\beta) = [X_i - \hat{g}_2(U_i)]\{Y_i - \hat{g}_{1\hat{G}}(U_i) - \beta^{\mathrm{T}}[X_i - \hat{g}_2(U_i)]\},
$$

其中 $\hat{g}_{1\hat{G}}(u)$ 和 $\hat{g}_2(u)$ 在 (2.3.5) 式中给出. 因此, β 的估计的经验对数似然比函数定义为

$$
\hat{l}_{\mathrm{E}}(\beta) = -2\max\left\{\sum_{i=1}^n \log(np_i)\middle| p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\hat{\eta}_i(\beta) = 0\right\}.
$$

由于 $\{\hat{\eta}_i(\beta_0), 1 \leqslant i \leqslant n\}$ 不是独立同分布的, 因此 $\hat{l}_{\mathrm{E}}(\beta_0)$ 不是渐近于标准 χ^2 变量. Wang 和 Li(2002) 在一些条件下证明了

$$
\hat{l}_{\mathrm{E}}(\beta_0) \xrightarrow{D} w_1\chi_{1,1}^2 + \cdots + w_p\chi_{1,p}^2, \tag{2.3.6}
$$

其中$\{\chi_{1,i}^2, 1\leqslant i\leqslant n\}$是自由度为 1 的独立 χ_1^2 随机变量, $\{w_i\}_{i=1}^p$ 是矩阵 $\Sigma_1^{-1}(\beta_0)\Sigma(\beta_0)$ 的特征根. 这里 $\Sigma(\beta_0)=\Sigma_1(\beta_0)-\Sigma_2(\beta_0)$,

$$\Sigma_1(\beta_0)=E\{[X_1-E(X_1|U_1)]^{\otimes 2}[Z_{1G}-\beta_0^{\mathrm{T}}X_1-g(U_1)]^2\},$$
$$\Sigma_2(\beta_0)=\int_{-\infty}^{\infty}M(u)M^{\mathrm{T}}(u)[1-F(u-)][1-\Delta\Lambda(u)]\mathrm{d}G(u),$$
$$M(u)=\frac{E\{[X_1-E(X_1|U_1)]Z_{1G}I(Z_1>z)\}}{[1-G(u)][1-F(u-)]},$$
$$\Lambda(u)=\int_{-\infty}^{u}\frac{1}{1-G(t-)}\mathrm{d}G(t).$$

此处 F 和 G 可以不连续. 如果给出未知权的估计, 那么 (2.3.6) 可以被用来构造 β_0 的置信域. 用 $\hat{F}(u)$ 表示 $F(u)$ 的 Kaplan-Meier 估计 (类似于 (2.3.3) 式). 定义 $\hat{\Sigma}(\beta_0)=\hat{\Sigma}_1(\beta_0)-\hat{\Sigma}_2(\beta_0)$,

$$\hat{\Sigma}_1(\beta_0)=\frac{1}{n}\sum_{i=1}^{n}\{[X_i-\hat{g}_2(U_i)]^{\otimes 2}[Z_{iG}-\hat{g}_{1\hat{G}}(U_i)-\beta_0^{\mathrm{T}}(X_i-\hat{g}_2(U_i))]^2\},$$
$$\hat{\Sigma}_2(\beta_0)=\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\hat{M}(Z_i)\hat{M}^{\mathrm{T}}(Z_i)[1-\Delta\hat{\Lambda}(Z_i-)],$$
$$\hat{M}(u)=\frac{n^{-1}\sum_{i=1}^{n}\{[X_i-\hat{g}_2(U_i)]Z_{i\hat{G}}I(Z_i>u)\}}{[1-\hat{G}(u)][1-\hat{F}(u-)]},$$
$$\hat{\Lambda}(u)=\frac{1}{n}\sum_{i=1}^{n}\frac{(1-\Delta_i)I(Z_i\leqslant u)}{[1-\hat{G}(u-)][1-\hat{F}(u-)]}.$$

设 $\hat{w}_1,\cdots,\hat{w}_p$ 是矩阵 $\hat{\Sigma}_1^{-1}(\hat{\beta}_{\mathrm{C}})\Sigma(\hat{\beta}_{\mathrm{C}})$ 的特征根, 其中 $\hat{\beta}_{\mathrm{C}}$ 在 (2.3.4) 式中定义. 用 c_α 表示给定 $\{(X_1,U_i,Z_i,\Delta_i),1\leqslant i\leqslant n\}$ 下 $\hat{w}_1\chi_{1,1}+\cdots+\hat{w}_p\chi_{1,p}$ 的条件分布的 $1-\alpha$ 分位数. 那么, β_0 的近似 $1-\alpha$ 置信域定义为

$$\mathcal{R}_{\mathrm{E}}(\beta)=\{\beta\,|\,\hat{l}_{\mathrm{E}}(\beta)\leqslant c_\alpha\}.$$

在实际中, 通过产生自由度为 1 的 χ^2 分布的 iid 样本 $\chi_{1,1}^2,\cdots,\chi_{1,p}^2$ 来计算分位数 c_α.

2. *调整的经验似然*

本节介绍构造 β_0 的置信域的另一个方法. 由 Rao 和 Scott(1981), $p\sum_{i=1}^{p}w_i\chi_{1,i}^2\Big/\sum_{i=1}^{p}w_i$ 的分布可以用 χ^2 分布逼近. 从而, 由 (2.3.6) 式可知, $p\hat{l}(\beta_0)/\mathrm{tr}(\hat{\Sigma}_1^{-1}(\beta_0)\hat{\Sigma}(\beta_0))$

的渐近分布可以用 χ^2 分布逼近. 然而, 这种逼近的精度依赖于诸 w_i 的值. 因此, 一个调整的经验似然比函数定义为

$$\hat{l}_{\mathrm{A}}(\beta) = r_n(\beta)\hat{l}_{\mathrm{E}}(\beta),$$

其中

$$r_n(\beta) = \frac{\mathrm{tr}(\hat{\Sigma}^{-1}(\beta)S_n(\beta))}{\mathrm{tr}(\hat{\Sigma}_1^{-1}(\beta)S_n(\beta))},$$

$$S_n(\beta) = \left[\sum_{i=1}^{n}\hat{\eta}_i(\beta)\right]\left[\sum_{i=1}^{n}\hat{\eta}_i(\beta)\right]^{\mathrm{T}}.$$

Wang 和 Li(2002) 在适当条件下证明了 $\hat{l}_{\mathrm{A}}(\beta_0)$ 以分布收敛于自由度为 p 的 χ^2 变量. 这个结果可以用来构造 β_0 的置信域:

$$\mathcal{R}_{\mathrm{A}}(\beta) = \{\beta\,|\hat{l}_{\mathrm{A}}(\beta) \leqslant \chi_p^2(1-\alpha)\}.$$

它也可以对假设 $H_0:\beta=\beta_0$ 进行检验.

2.4 测量误差数据分析

本节主要讨论协变量具有测量误差的部分线性 EV(errors in variables) 模型, 其形式为

$$\begin{cases} Y_i = \beta_0^{\mathrm{T}}X_i + g(U_i) + \varepsilon_i, \\ V_i = X_i + \eta_i, \quad i = 1, 2, \cdots, n, \end{cases} \tag{2.4.1}$$

其中 β_0 是 p 维参数向量, $g(\cdot)$ 是定义在 $[0,1]$ 上的未知函数. ε_i 是随机误差, 且几乎处处有 $E(\varepsilon_i|X_i,U_i) = 0$, X_i 是来自 $X \in \mathbf{R}^p$ 的不可观测的随机向量, $\{(V_i,U_i,Y_i), 1 \leqslant i \leqslant n\}$ 是来自 (V,U,Y) 的 iid 样本, 测量误差 $\{\eta_i, 1 \leqslant i \leqslant n\}$ 相互独立且与 η 同分布, 具有均值 0 和协方差阵 Σ_η, 且独立于 $\{(X_i,U_i,Y_i), 1 \leqslant i \leqslant n\}$. 考虑到模型的可识别性, 假定 Σ_η 已知. 我们的主要兴趣集中在参数分量和非参数分量的统计推断上, 其中包括估计方法和经验似然.

2.4.1 校正的最小二乘估计

正如 Liang 等 (1999) 指出, 如果忽略测量误差而直接用 V_i 代替 (2.2.3) 式中的 X_i, 所得到的估计对 β_0 是不相合的. 众所周知, 在线性模型中, 由测量误差引起的不相合性可以应用所谓 "校正衰减 (correction attenuation)" 来克服. 对模型 (2.4.1), Liang 等 (1999) 建议使用估计量

$$\hat{\beta}_{\mathrm{CA}} = [(\boldsymbol{V}-\hat{\boldsymbol{g}}_3)^{\mathrm{T}}(\boldsymbol{V}-\hat{\boldsymbol{g}}_3) - n\Sigma_\eta]^{-1}(\boldsymbol{V}-\hat{\boldsymbol{g}}_3)(\boldsymbol{Y}-\hat{\boldsymbol{g}}_1), \tag{2.4.2}$$

其中 $\boldsymbol{V}=(V_1,\cdots,V_n)^{\mathrm{T}}$, $\boldsymbol{Y}=(Y_1,\cdots,Y_n)^{\mathrm{T}}$, $\hat{\boldsymbol{g}}_\nu=(\hat{g}_\nu(U_1),\cdots,\hat{g}_\nu(U_n))^{\mathrm{T}}$, $\nu=1,3$,

$$\hat{g}_1(u)=\sum_{i=1}^n W_{ni}(u)Y_i,\quad \hat{g}_3(u)=\sum_{i=1}^n W_{ni}(u)V_i. \tag{2.4.3}$$

此处 $\hat{g}_1(u)$ 和 $\hat{g}_3(u)$ 分别是 $g_1(u)=E(Y|U=u)$ 和 $g_3(u)=E(V|U=u)$ 的估计量, $\{W_{ni}(u),1\leqslant i\leqslant n\}$ 可取任何权函数, 如核权或局部多项式权. 注意到

$$g(u)=E[(Y-\beta_0X)|U=u]=E[(Y-\beta_0^{\mathrm{T}}V)|U=u].$$

因此, $g(u)$ 的最终估计量定义为

$$\hat{g}_{\mathrm{CA}}(u)\equiv\hat{g}(u;\hat{\beta}_{\mathrm{CA}})=\sum_{i=1}^n W_{ni}(u)(Y_i-\hat{\beta}_{\mathrm{CA}}^{\mathrm{T}}V_i). \tag{2.4.4}$$

由 (2.4.2) 式定义的估计量可以按如下理解. 注意到

$$E[Y-\beta_0^{\mathrm{T}}X-g(U)]^2=E[Y-\beta_0^{\mathrm{T}}V-g(U)]^2-\beta_0^{\mathrm{T}}\Sigma_\eta\beta_0.$$

因此, 利用 $\hat{g}(\cdot;\beta)$ 可定义 β_0 的估计量 $\hat{\beta}_{\mathrm{CA}}$ 为极小化目标函数

$$S(\beta)=\frac{1}{n}\sum_{i=1}^n[Y_i-\beta^{\mathrm{T}}V_i-\hat{g}(U_i;\beta)]^2-\beta^{\mathrm{T}}\Sigma_\eta\beta \tag{2.4.5}$$

的解. (2.4.5) 式右端第二项 $-\beta^{\mathrm{T}}\Sigma_\eta\beta$ 是一个负的惩罚运算, 它的作用是削减由测量误差带来的影响. 也就是说, 当它是一个标量时, 为了在绝对值中低估 β, 人们必须取更大的 β 来校正衰减, 而不是进一步皱缩它接近于 0. 第二项 $-\beta^{\mathrm{T}}\Sigma_\eta\beta$ 起一个校正的作用, 因此称 $\hat{\beta}_{\mathrm{CA}}$ 为校正的最小二乘估计.

在某些情况下, 合理的假定是模型误差 ε_i 是同方差的且具有方差 $\sigma^2>0$. 在这种情况下, 因为 $\sigma^2=E[Y_i-\beta_0^{\mathrm{T}}V_i-g(U_i)]^2-\beta_0\Sigma_\eta\beta_0$. 因此定义

$$\hat{\sigma}_{\mathrm{CA}}^2=\frac{1}{n}\sum_{i=1}^n[Y_i-\hat{\beta}_{\mathrm{CA}}^{\mathrm{T}}V_i-\hat{g}_{\mathrm{CA}}(U_i)]^2-\hat{\beta}_{\mathrm{CA}}\Sigma_\eta\hat{\beta}_{\mathrm{CA}}$$

作为 σ^2 的估计量.

Liang 等 (1999) 在一般的权函数下研究了 $\hat{\beta}_{\mathrm{CA}}$, $\hat{\sigma}_{\mathrm{CA}}^2$ 和 $\hat{g}_{\mathrm{CA}}(u)$ 的大样本性质. 关于权函数的条件如下.

条件 2.4.1 权函数 $W_{ni}(u)$ 满足:

$$\max_{1\leqslant i\leqslant n}\sum_{j=1}^n W_{nj}(U_i)=O_P(1),$$

$$\max_{1\leqslant i,j\leqslant n} W_{nj}(U_i)=O_P(b_n),$$

$$\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n} W_{nj}(U_i)I(|U_j-U_i|>c_n)=O_P(c_n),$$

其中 b_n 和 c_n 为趋于 0 的正的常数序列. 例如, 取 Gasser-Müller(1979) 权函数

$$W_{ni}(u)=\int_{s_{i-1}}^{s_i} K_h(t-u)\mathrm{d}t,\quad i=1,\cdots,n,$$

其中 $s_0=0$, $s_n=1$, $s_i=(U_{(i)}+U_{(i+1)})/2$, $1\leqslant i\leqslant n-1$, $\{U_{(i)},1\leqslant i\leqslant n\}$ 为 $\{U_i,1\leqslant i\leqslant n\}$ 的次序统计量, $h=h_n$ 是一列收敛于 0 的带宽, $K_h(u)=h^{-1}K(u/h)$, $K(u)$ 为 $[-1,1]$ 上的有界且对称的概率密度函数. 可以验证 Gasser-Müller(1979) 权函数满足条件 2.4.1.

Liang 等 (1999) 给出的下面两个定理陈述了 $\hat\beta_{\mathrm{CA}}$ 和 $\hat\sigma^2_{\mathrm{CA}}$ 的渐近正态性.

定理 2.4.1　设条件 2.4.1 成立, 其中 $b_n=n^{-4/5}$, $c_n=n^{-1/5}$, 又设 $\sup\limits_{0\leqslant u\leqslant 1} E(\|X\|^4|U=u)<\infty$, $E(\varepsilon^4+\|\eta\|^4)<\infty$, $B=[X-E(X|U)]^{\otimes 2}$ 为正定矩阵, 则

$$\sqrt{n}(\hat\beta_{\mathrm{CA}}-\beta_0)\xrightarrow{D} N(0,B^{-1}\Gamma B^{-1}),$$

其中 $\Gamma=E\{(\varepsilon-\beta_0^{\mathrm T}\eta)[X-E(X|U)]\}^{\otimes 2}+E[(\eta\eta^{\mathrm T}-\Sigma_\eta)\beta_0]^{\otimes 2}+E(\varepsilon^2\eta\eta^{\mathrm T})$. 这里 $A^{\otimes 2}=AA^{\mathrm T}$, ε 与 ε_i 同分布. 如果 ε_i 为同方差的且独立于 (X_i,U_i), 那么 $\Gamma=E(\varepsilon-\beta_0^{\mathrm T}\eta)^2B+E[(\eta\eta^{\mathrm T}-\Sigma_\eta)\beta_0]^{\otimes 2}+\sigma^2\Sigma_\eta$.

定理 2.4.2　假设定理 2.4.1 的条件成立, 又设 ε_i 为同方差的, 其方差为 σ^2, 且独立于 (X_i,U_i), 则

$$\sqrt{n}(\hat\sigma^2_{\mathrm{CA}}-\sigma^2)\xrightarrow{D} N(0,\gamma^2),$$

其中 $\gamma^2=E\{(\varepsilon-\beta_0^{\mathrm T}\eta)^2-(\beta_0^{\mathrm T}\Sigma_\eta\beta_0)\}^2$.

估计 $\hat\beta_{\mathrm{CA}}$ 的渐近方差相对容易. B 和 Γ 的相合估计分别为

$$\hat B=\frac{1}{n-p}\sum_{i=1}^{n}[V_i-\hat g_3(U_i)]^{\otimes 2}-\Sigma_\eta,$$

$$\hat\Gamma=\frac{1}{n}\sum_{i=1}^{n}\left\{[V_i-\hat g_3(U_i)][Y_i-\hat\beta_{\mathrm{CA}}^{\mathrm T}V_i-\hat g_{\mathrm{CA}}(U_i)]+\Sigma_\eta\hat\beta_{\mathrm{CA}}\right\}^{\otimes 2}.$$

在同方差的情况, 即 ε_i 独立于 (X_i,U_i,η_i) 且具有方差 σ^2, η 为正态分布, Γ 的相合估计为

$$(\hat\sigma^2_{\mathrm{CA}}+\hat\beta_{\mathrm{CA}}\Sigma_\eta\hat\beta_{\mathrm{CA}})\hat B+E[(\eta\eta^{\mathrm T}-\Sigma_\eta)\hat\beta_{\mathrm{CA}}]^{\otimes 2}+\hat\sigma^2_{\mathrm{CA}}\Sigma_\eta.$$

下面的定理给出了 $\hat g_{\mathrm{CA}}(u)$ 的渐近偏差和渐近方差.

定理 2.4.3 假设定理 2.4.1 的条件成立, 又设 $W_{ni}(u)$ 满足 1 阶 Lipschitz 条件, $i=1,\cdots,n$, 则对固定的 u, 估计量 $\hat{g}_{\mathrm{CA}}(u)$ 的渐近偏差和渐近方差分别为 $\sum_{i=1}^{n} W_{ni}(u)g(U_i)-g(u)$ 和 $\sum_{i=1}^{n} W_{ni}^2(u)(\beta_0^{\mathrm{T}}\Sigma_\eta\beta_0+\sigma^2)$. 特别地, 对核估计, 二者的阶都为 $O(n^{-2/5})$.

2.4.2 广义最小二乘估计

在古典的函数模型中 (Kendall and Stuart, 1992), 可以复制 Σ_η 的一个估计. 因此可以假定 Σ_η 与 σ^2 的比已知. 不失一般性, 设 Σ_η/σ^2 为单位矩阵. 因此, 在 ε_i 为同方差的情况下,

$$\frac{E[Y-\beta_0^{\mathrm{T}}V-g(U)]^2}{1+||\beta_0||^2}=\sigma^2.$$

于是, 按普通的最小二乘法, 通过求解关于 β 的极小化问题

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i-\beta^{\mathrm{T}}V_i-\hat{g}(U_i;\beta)}{\sqrt{1+||\beta||^2}}\right)^2=\min!, \tag{2.4.6}$$

可以得到 β_0 的另一个估计, 记为 $\hat{\beta}_{\mathrm{G}}$, 并称之为广义最小二乘估计. 由此可定义 $g(u)$ 和 σ^2 的估计量:

$$\hat{g}_{\mathrm{G}}(u)=\hat{g}_1(u)-\hat{\beta}_{\mathrm{G}}^{\mathrm{T}}\hat{g}_3(U_i),$$

$$\hat{\sigma}_{\mathrm{G}}^2=\frac{1}{n}\sum_{i=1}^{n}\frac{\{[Y_i-\hat{g}_1(U_i)]-\hat{\beta}_{\mathrm{G}}^{\mathrm{T}}[V_i-\hat{g}_3(U_i)]\}^2}{1+||\hat{\beta}_{\mathrm{G}}||^2}.$$

崔恒建 (1995) 证明了估计量 $\hat{\beta}_{\mathrm{G}}$ 和 $\hat{\sigma}_{\mathrm{G}}^2$ 的相合性和渐近正态性, 并给出了 $\hat{g}_{\mathrm{G}}(u)$ 的收敛速度. 当 ε_i 与 (X_i,U_i) 独立时, $\hat{\beta}_{\mathrm{G}}$ 的渐近方差为

$$B^{-1}\left\{(1+||\beta_0||^2)^2\sigma^2B+\frac{E[(\varepsilon-\beta_0^{\mathrm{T}}\eta)^2\Gamma_1\Gamma_1^{\mathrm{T}}]}{1+||\beta_0||^2}\right\}B^{-1},$$

其中 $\Gamma_1=(1+||\beta_0||^2)\eta+(\varepsilon-\beta_0^{\mathrm{T}}\eta)\beta_0$.

当 Σ_η 未知时, Liang 等 (1999) 利用无偏的矩估计方法构造了它的相合估计, 同时也给出了 β_0 的一个修正的估计量. 这里不再赘述.

2.4.3 分位数回归估计

为构造 β_0 的 τ 分位数回归估计, 首先定义 τ 分位数损失函数

$$\rho_\tau(t)=\tau\max\{t,0\}+(1-\tau)\max\{-t,0\}. \tag{2.4.7}$$

对随机变量 X 和 U 提出下列假定:

条件 2.4.2　对一切 $u \in [0,1]$, $E(X|U=u)=0$.

利用 (2.4.4) 式的 $\hat{g}(\cdot;\beta)$ 可定义 β_0 的估计量 $\hat{\beta}_\tau$ 为极小化目标函数

$$Q(\beta)=\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(\frac{Y_i-\beta^{\mathrm{T}}V_i-\hat{g}(U_i;\beta)}{\sqrt{1+||\beta||^2}}\right) \tag{2.4.8}$$

的解.

注意到损失函数 $\rho_\tau(t)$ 除 0 点外是处处可微的. 目标函数 $Q(\beta)$ 在 $\hat{\beta}_\tau$ 处的方向导数是非负的. 因此可以推出, 在 $\beta=\hat{\beta}_\tau$ 处,

$$\begin{aligned}&\sum_{i=1}^{n}\left(\frac{V_i-\hat{g}_3(U_i)-[Y_i-\beta^{\mathrm{T}}V_i-\hat{g}(U_i;\beta)]\beta}{1+||\beta||^2}\right)\psi_\tau\left(\frac{Y_i-\beta^{\mathrm{T}}V_i-\hat{g}(U_i;\beta)}{\sqrt{1+||\beta||^2}}\right)\\&=O\left(\sum_{i\in h}[V_i-g_3(U_i)]\right),\quad \text{a.s.,}\end{aligned}$$

其中 $\psi_\tau(t)$ 为 $\rho_\tau(t)$ 的方向导数, $\{i\in h\}$ 为零残差指标集. 即解 $\hat{\beta}_\tau$ 不完全满足估计方程, 但假如 (V,Y) 的分布连续, 就对任何线性拟合的零残差的数目以概率 1 小于或等于 $p+1$ 而论, 它的确非常接近.

分位数回归估计被看成一种特别类型的 M 估计, 许多作者从稳健性角度出发研究了它们的统计性质. 这里考虑的分位数回归估计问题在方法上不同于现有文献的 M 估计, 例如, 这里不需要诸如正态的中心误差模型. 分位数对非正态模型是特别有趣的. 在相合性框架下, 分位数回归估计具有渐近正态分布. 用 $f(\cdot)$ 表示 ε 的密度函数. He 和 Liang(2000) 证明了下列结果.

定理 2.4.4　设 $g(u)$ 满足 1 阶 Lipschitz 条件, $B=E(XX^{\mathrm{T}})$ 为正定矩阵, 误差 $(\varepsilon,\eta^{\mathrm{T}})$ 的分布为球对称的, q_τ 为 $E[\rho_\tau(\varepsilon-q)]=0$ 的唯一解, $f(q_\tau)>0$, $\lim\limits_{e\to 0}[f(e+q_\tau)-f(q_\tau)]=0$, 且对某个常数 $0<\delta<1$, $E(|\varepsilon|^{2+\delta})<\infty$, $E(||X||^{2+\delta})<\infty$. 又设条件 2.4.1 和条件 2.4.2 成立, 其中 $b_n=n^{-(3-\delta_1)/4}$, $c_n=\log n/(nb_n)$, $\delta_1=\delta/(2+\delta)$. 则 $\hat{\beta}_\tau$ 是 β_0 的相合估计, 且

$$\sqrt{n}(\hat{\beta}_\tau-\beta_0)\xrightarrow{D}N(0,\Sigma),$$

其中

$$\begin{aligned}\Sigma&=f^{-2}(q_\tau)(1+||\beta_0||^2)B^{-1}AB^{-1},\\A&=\tau(1-\tau)B+\mathrm{cov}(\psi_\tau(\xi)(\eta+\xi\beta_0/\sqrt{1+||\beta_0||^2})),\\\xi&=(\varepsilon-\beta_0^{\mathrm{T}}\eta)/\sqrt{1+||\beta_0||^2}-q_\tau.\end{aligned}$$

如果没有测量误差, 那么 $A=\tau(1-\tau)B$. 如果 $(\varepsilon,\eta^{\mathrm{T}})$ 是多元正态的, 那么 A 的表达式可简化为

$$A=\tau(1-\tau)B+E[\eta\eta^{\mathrm{T}}\psi_\tau^2(\xi)]-E[\xi^2\psi_\tau^2(\xi)]\beta_0\beta_0^{\mathrm{T}}/(1+\|\beta_0\|^2).$$

He 和 Liang(2000) 选取核权函数

$$W_{ni}(u)=K_h(U_i-u)\Big/\sum_{i=1}^n K_h(U_i-u),\quad i=1,\cdots,n, \tag{2.4.9}$$

并在适当的条件下证明了条件 2.4.1 成立. 即有下列结果.

命题 2.4.1 设核 $K(u)$ 为 $[-1,1]$ 上的有界且对称的概率密度函数, 带宽 $h=c/(nb_n)$, c 为某个正的常数, 存在正的常数 c_1 和 c_2 使得 $c_1/n\leqslant\min\{|U_i-U_{i-1}|\}\leqslant\max\{|U_i-U_{i-1}|\}\leqslant c_2/n$. 则由 (2.4.9) 式定义的权函数满足条件 2.4.1.

简单运算即可证得命题 2.4.1, 这里省略其证明. 一个特别的例子是使用 Nadaraya-Watson 核 $K(u)=(15/16)(1-u^2)^2I(|u|\leqslant1)$. 建议在定理 2.4.4 中使用 $h=cn^{-(1+\delta_1)/4}$, $\delta_1>0$ 为某个常数. 正如 He 和 Liang(2000) 指出的那样, 主要目的是估计 β_0, 对带宽 h 的选择不像在完全的非参数估计中那样至关重要.

2.4.4 经验似然

首先假定 $g_1(u)=E(Y|U=u)$ 和 $g_3(u)=E(V|U=u)$ 是已知的. 对任何参数向量 β, 构造辅助随机变量

$$Z_i(\beta)=[V_i-g_3(U_i)]\{Y_i-g_1(U_i)-[V_i-g_3(U_i)]^{\mathrm{T}}\beta\}+\Sigma_\eta\beta. \tag{2.4.10}$$

上式右端第二项是为了避免测量误差所造成的影响而进行的纠偏. 容易验证: 当 $\beta=\beta_0$ 时, 有 $E[Z_i(\beta_0)]=0$. 因而检验 $\beta=\beta_0$ 等价于检验 $E[Z_i(\beta_0)]=0$, $i=1,\cdots,n$. 根据文献 (Owen, 1991), 可以采用经验似然方法. 然而, 利用 $Z_i(\beta_0)$ 所构造的经验对数似然比虽然渐近于 χ^2 变量, 但 $Z_i(\beta_0)$ 中含有未知量 $g_1(u)$ 和 $g_3(u)$, 其结果不能被直接使用. 因此, 我们需要用 (2.4.3) 所定义的估计量 $\hat g_1(u)$ 和 $\hat g_3(u)$ 代替 $g_1(u)$ 和 $g_3(u)$. 因此, 引入辅助随机变量

$$\hat Z_i(\beta)=[V_i-\hat g_3(U_i)]\{Y_i-\hat g_1(U_i)-\beta^{\mathrm{T}}[V_i-\hat g_3(U_i)]\}+\Sigma_\eta\beta.$$

β_0 的估计的经验对数似然比函数定义为

$$\hat l_{\mathrm{EV}}(\beta)=-2\max\left\{\sum_{i=1}^n\log(np_i)\Big|\,p_i\geqslant0,\sum_{i=1}^n p_i=1,\sum_{i=1}^n p_i\hat Z_i(\beta)=0\right\}.$$

可以证明: 在适当的条件下 $\hat l_{\mathrm{EV}}(\beta_0)$ 以分布收敛于自由度为 p 的 χ^2 变量. 这个结果可以用来构造 β_0 的置信域:

$$\{\beta \,|\hat{l}_{\rm EV}(\beta) \leqslant \chi_p^2(1-\alpha)\}.$$

极大化 $\{-\hat{l}_{\rm EV}(\beta)\}$ 可以得到 β_0 的极大经验似然估计量 $\hat{\beta}_{\rm EV}$. 可以证明 $\hat{\beta}_{\rm EV}$ 如同 (2.4.2) 式所定义的估计量 $\hat{\beta}_{\rm CA}$, 且二者具有相同的渐近方差. 需要说明的是, 刘强等 (2009) 研究了删失数据下部分线性 EV 模型中参数的经验似然置信域的构造问题.

2.5 缺失数据分析

本节主要讨论在响应变量和协变量缺失下部分线性模型中兴趣参数的估计方法和置信域构造问题, 内容主要取材于文献 (Wang et al., 2004; Liang et al., 2007; Wang, 2009; Yang et al., 2009; Xue et al., 2011).

2.5.1 缺失响应下回归系数的估计

在实际问题中, 在容量为 n 的样本中一些 Y 的值可能缺失, 但 X 和 U 可以被完全观测. 换句话说, 数据是由来自 (X,U,Y,δ) 的不完全样本 $\{(X_i,U_i,Y_i,\delta_i), 1\leqslant i\leqslant n\}$ 组成, 其中诸 X_i 和 U_i 可被观测, 而如果 Y_i 缺失, 则 $\delta_i=0$, 否则 $\delta_i=1$. 本节假定 Y 随机缺失 (missing at random, MAR), 即

$$P(\delta=1|X,U,Y)=P(\delta=1|X,U)\equiv p(X,U).$$

上述假定意味着 δ 与 Y 条件独立. 设 $\{(X_i,U_i,Y_i,\delta_i), 1\leqslant i\leqslant n\}$ 为 iid 样本. 假定这个数据集可以由 (2.1.1) 式建模, 其中 $\{\varepsilon_i, 1\leqslant i\leqslant n\}$ 和 $\{\delta_i, 1\leqslant i\leqslant n\}$ 相互独立.

1. **完全情况数据 (complete-case data) 下回归系数的估计**

在 (2.1.1) 式左右两端分别乘以 δ_i 可得

$$\delta_i Y_i=\delta_i\beta_0^{\rm T}X_i+\delta_i g(U_i)+\delta_i\varepsilon_i,\quad i=1,\cdots,n,$$

再将上式两端取关于 U_i 的条件期望, 得

$$E(\delta_iY_i|U_i=u)=\beta_0^{\rm T}E(\delta_iX_i|U_i=u)+E(\delta_i|U_i=u)g(u),$$

由此可得

$$g(u)=\mu_1(u)-\beta_0^{\rm T}\mu_2(u), \tag{2.5.1}$$

其中

$$\mu_1(u)=\frac{E(\delta Y|U=u)}{E(\delta|U=u)},\quad \mu_2(u)=\frac{E(\delta X|U=u)}{E(\delta|U=u)}.$$

由 (2.1.1) 和 (2.5.1) 式可以推出

$$[Y_i - \mu_1(U_i)] = \beta_0^{\mathrm{T}}[X_i - \mu_2(U_i)] + \varepsilon_i, \quad i = 1, \cdots, n.$$

因此, 引入辅助随机变量

$$\eta_{i,\mathrm{C}}(\beta_0) = \delta_i\{[Y_i - \mu_1(U_i)] - \beta_0^{\mathrm{T}}[X_i - \mu_2(U_i)]\}[X_i - \mu_2(U_i)].$$

在 MAR 下, $E(\eta_{i,\mathrm{C}}(\beta_0)) = 0$. 使用这个信息, 可以定义 β_0 的经验对数似然比函数 $R_{\mathrm{C}}(\beta_0)$. 利用 Owen(1990) 的证明方法, 可以推出 $R_{\mathrm{C}}(\beta_0)$ 近似地服从自由度为 p 的 χ^2 分布. 不幸的是, $\mu_1(\cdot)$ 和 $\mu_2(\cdot)$ 通常是未知的, 因此, $R_{\mathrm{C}}(\beta_0)$ 不能直接用来对 β_0 作统计推断. 解决这个问题的一个自然思想是用它们的估计代替这些未知量. 这里使用 Wang 等 (2004) 定义的核估计. 设 $K(\cdot)$ 是 Borel 可测核函数, $h = h_n$ 是收敛于 0 正的带宽序列, 那么定义 $\mu_1(u)$ 和 $\mu_2(u)$ 的估计量:

$$\hat{\mu}_1(u) = \sum_{i=1}^{n} \Omega_{ni}(u) Y_i, \quad \hat{\mu}_2(u) = \sum_{i=1}^{n} \Omega_{ni}(u) X_i, \tag{2.5.2}$$

其中

$$\Omega_{ni}(u) = \frac{\delta_i K_h(U_i - u)}{\sum_{j=1}^{n} \delta_j K_h(U_j - u)}, \tag{2.5.3}$$

且 $K_h(\cdot) = h^{-1}K(\cdot/h)$. $\eta_{i,\mathrm{C}}(\beta_0)$ 的估计 $\hat{\eta}_{i,\mathrm{C}}(\beta_0)$ 可以用 $\hat{\mu}_1(U_i)$ 和 $\hat{\mu}_2(U_i)$ 代替 $\eta_{i,\mathrm{C}}(\beta)$ 中的 $\mu_2(U_i)$ 和 $\mu_1(U_i)$ 而得到, 即

$$\hat{\eta}_{i,\mathrm{C}}(\beta_0) = \delta_i \check{X}_i(\check{Y}_i - \beta_0^{\mathrm{T}}\check{X}_i),$$

其中 $\check{Y}_i = Y_i - \hat{\mu}_1(U_i)$, $\check{X}_i = X_i - \hat{\mu}_2(U_i)$. 因此, β_0 的具有完全情况数据的经验对数似然比函数定义为

$$\hat{R}_{\mathrm{C}}(\beta_0) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\middle| p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\hat{\eta}_{i,\mathrm{C}}(\beta_0) = 0\right\}.$$

为了获得 $\hat{R}_{\mathrm{C}}(\beta_0)$ 的渐近性质, 首先给出下列一组条件.

(C2.5.1) $q(u)$ 和 $f(u)$ 在 $(0,1)$ 上具有有界连续的导数, 且 $f(u)$ 在 $[0,1]$ 上有界远离 0, 即 $0 < \inf_{u\in[0,1]} f(u) \leqslant \sup_{u\in[0,1]} f(u) < \infty$, 其中 $f(u)$ 为 U 的密度函数, $q(u) = E(\delta|U = u)$.

(C2.5.2) 函数 $g(u)$, $\mu_1(u)$ 和 $\mu_{2s}(u)$ 在 $(0,1)$ 上具有有界连续的 2 阶导数, 其中 $\mu_{2s}(u)$ 是 $\mu_2(u)$ 的第 s 个分量.

(C2.5.3) $\sup_{x,u} E(\varepsilon^2|X = x, U = u) < \infty$, $\sup_{u} E(||X||^2|U = u) < \infty$.

(C2.5.4) 核 $K(u)$ 是具有支撑 $[-1,1]$ 的对称且有界的概率密度函数.

(C2.5.5) 带宽 h 满足 $h=h_0n^{-1/5}$, 其中 $h_0>0$ 为某个常数.

(C2.5.6) 矩阵 $A=E\{p(X,U)[X-\mu_2(U)]^{\otimes 2}\}$ 是正定的, 其中 $H^{\otimes 2}=HH^{\mathrm{T}}$.

注 2.5.1　光滑条件 (C2.5.1) 和 (C2.5.2) 是非参数问题的标准条件. 条件 (C2.5.3) 和 (C2.5.6) 是得到 $\hat{\beta}_{\mathrm{C}}$ 的渐近正态性所必需的条件. 条件 (C2.5.4) 是对核函数的通常假定. 条件 (C2.5.5) 放宽了 Wang 等 (2004) 对带宽 h 所施加的条件.

下面的定理给出了 $\hat{R}_{\mathrm{C}}(\beta_0)$ 的渐近性质.

定理 2.5.1　设条件 (C2.5.1)~(C2.5.5) 成立, 则 $\hat{R}_{\mathrm{C}}(\beta_0)\xrightarrow{D}\chi_p^2$.

利用定理 2.5.1, 可以得到 β_0 的近似 $1-\alpha$ 置信域

$$\left\{\beta|\ \hat{R}_{\mathrm{C}}(\beta)\leqslant\chi_p^2(1-\alpha)\right\}.$$

定理 2.5.1 也可以用来检验假设 $H_0:\beta=\beta_0$, 即, 如果 $\hat{R}_{\mathrm{C}}(\beta_0)>\chi_p^2(1-\alpha)$, 则在水平 α 下拒绝 H_0.

我们可以极小化 $\hat{R}_{\mathrm{C}}(\beta)$ 得到 β_0 的完全情况下的估计, 记为 $\hat{\beta}_{\mathrm{C}}$. 由于 $\hat{R}_{\mathrm{C}}(\beta)\geqslant 0$, 于是容易看出, 当 $p_i=1/n$ 时 $\hat{R}_{\mathrm{C}}(\beta)$ 在 $\hat{\beta}_{\mathrm{C}}$ 达到最小值. 这就推出 $\hat{\beta}_{\mathrm{C}}$ 是估计方程 $\sum\limits_{i=1}^{n}\hat{\eta}_{i,\mathrm{C}}(\beta)=0$ 的解, 即

$$\hat{\beta}_{\mathrm{C}}=\left(\sum_{i=1}^{n}\delta_i\breve{X}_i\breve{X}_i^{\mathrm{T}}\right)^{-1}\sum_{i=1}^{n}\delta_i\breve{X}_i\breve{Y}_i. \tag{2.5.4}$$

显然, 估计量 $\hat{\beta}_{\mathrm{C}}$ 就是 β_0 的最小二乘估计. 下列定理给出了 $\hat{\beta}_{\mathrm{C}}$ 的渐近正态性.

定理 2.5.2　假设条件 (C2.5.1)~(C2.5.6) 成立, 则

$$\sqrt{n}(\hat{\beta}_{\mathrm{C}}-\beta_0)\xrightarrow{D}N(0,\Sigma),$$

其中 $\Sigma=A^{-1}BA^{-1}$, A 在条件 (C2.5.6) 中定义, $B=E\{p(X,U)[X-\mu_2(U)]^{\otimes 2}\varepsilon^2\}$, $\varepsilon=Y-\beta_0^{\mathrm{T}}X-g(U)$.

为应用定理 2.5.2 构造 β_0 的置信域, 我们使用 $\hat{\Sigma}=\hat{A}^{-1}\hat{B}\hat{A}^{-1}$ 作为 Σ 相合估计, 其中 $\hat{A}$ 和 $\hat{B}$ 分别定义为

$$\hat{A}=\frac{1}{n}\sum_{i=1}^{n}\delta_i\breve{X}_i\breve{X}_i^{\mathrm{T}},\quad \hat{B}=\frac{1}{n}\sum_{i=1}^{n}\delta_i\breve{X}_i\breve{X}_i^{\mathrm{T}}(\breve{Y}_i-\breve{X}_i^{\mathrm{T}}\hat{\beta}_{\mathrm{C}})^2.$$

因此可得

$$\hat{\Sigma}^{-1/2}\sqrt{n}(\hat{\beta}_{\mathrm{C}}-\beta_0)\xrightarrow{D}N(1,I_p), \tag{2.5.5}$$

$$\left(\hat{\beta}_{\mathrm{C}}-\beta_0\right)^{\mathrm{T}} n\hat{\Sigma}^{-1}\left(\hat{\beta}_{\mathrm{C}}-\beta_0\right) \xrightarrow{D} \chi_p^2. \tag{2.5.6}$$

使用 (2.5.5) 和 (2.5.6) 可以构造 β_0 的大样本置信区间或置信域.

2. 具有借补值的回归系数的估计

具有完全情况数据的经验似然不能完全探究包含在数据中的信息. 当缺失值较多时, 由于丢弃了缺失的数据, 置信域的精度会实质性降低. 这里介绍一个借补方法来处理缺失数据问题. 按照线性回归借补的思想, 如果 Y_i 缺失, 就用 $\hat{\beta}_{\mathrm{C}}^{\mathrm{T}}\check{X}_i$ 填补 $\check{Y}_i$, 得到 $\check{Y}_i$ 的借补值, 即

$$\check{Y}_i=\frac{\delta_i\check{Y}_i}{\hat{p}(X_i,U_i)}+\left(1-\frac{\delta_i}{\hat{p}(X_i,U_i)}\right)\hat{\beta}_{\mathrm{C}}^{\mathrm{T}}\check{X}_i,$$

其中

$$\hat{p}(x,u)=\frac{\sum_{i=1}^{n}\delta_i\widetilde{K}((X_i-x)/b,(U_i-u)/b)}{\max\left\{1,\sum_{i=1}^{n}\widetilde{K}((X_i-x)/b,(U_i-u)/b)\right\}}, \tag{2.5.7}$$

$b=b_n$ 是趋于 0 的正的常数序列, $\widetilde{K}(\cdot)$ 是定义在 $\mathbf{R}^{p+1}$ 上的核函数. 于是可将 $\{(\check{X}_i,\check{Y}_i),1\leqslant i\leqslant n\}$ 看作完整数据, 从而引入辅助随机变量

$$\hat{\eta}_{i,\mathrm{I}}(\beta)=\check{X}_i(\check{Y}_i-\beta^{\mathrm{T}}\check{X}_i). \tag{2.5.8}$$

因此, 定义基于借补值的经验对数似然比函数

$$\hat{R}_{\mathrm{I}}(\beta)=-2\max\left\{\sum_{i=1}^{n}\log(np_i)\middle|\,p_i\geqslant 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\hat{\eta}_{i,\mathrm{I}}(\beta)=0\right\}.$$

此经验似然比函数比完全情况的经验似然比更合理, 因为它充分使用了包含在数据中的信息. 模拟结果证实了这个事实: 基于借补值的经验似然比完全情况的经验似然有更短的置信区间.

为给出 $\hat{R}_{\mathrm{I}}(\beta_0)$ 的渐近正性质, 除条件 (C2.5.1)~(C2.5.6) 外, 再给出下列条件.

(C2.5.7) $p(x,u)$ 和 $s(x,u)$ 具有 r 阶有界偏导数, 其中 $r>(p+1)/2$, $s(x,u)$ 表示 (X,U) 的密度函数; $\inf\limits_{x,u}p(x,u)>0$.

(C2.5.8) $\widetilde{K}(u)$ 是 r 阶核函数, r 在条件 (C2.5.7) 中定义; 存在正的常数 C_1,C_2 和 ρ 使得 $C_1I[||u||\leqslant\rho]\leqslant\widetilde{K}(u)\leqslant C_2I[||u||\leqslant\rho]$.

(C2.5.9) $P(||(X^{\mathrm{T}},U)^{\mathrm{T}}||>M_n)=o(n^{-1/2})$, 其中 $0<M_n\to\infty$, $n\to\infty$.

(C2.5.10) $nb^{2(p+1)}M_n^{-2(p+1)} \to \infty$, $nb^{4r} \to 0$, 其中 r 在条件 (C2.5.7) 中定义.

条件 (C2.5.7) 和 (2.5.8) 是非参数问题中常用的假定. 条件 (C2.5.9) 是为避免边界问题而通常使用的条件. 对下列情况, 条件 (C2.5.9) 满足: X 服从标准正态分布或指数分布, U 服从 $[0,1]$ 上的均匀分布, 且 $M_n = 4\ln n$. 条件 (C2.5.10) 放宽了 Wang 等 (2004) 对 b 所施加的条件.

定理 2.5.3 假设条件 (C2.5.1)~(C2.5.10) 成立, 则 $\hat{R}_{\rm I}(\beta_0) \xrightarrow{D} \chi_p^2$.

使用定理 2.5.3, 可以得到 β_0 的近似 $1-\alpha$ 置信域

$$\left\{\beta \middle| \hat{R}_{\rm I}(\beta) \leqslant \chi_p^2(1-\alpha)\right\}. \tag{2.5.9}$$

当 X 的维数较高时, 可以考虑 $p(x,u)$ 的参数或半参数模型. 假设我们的模型是 $p_\gamma(x,u)$, 其中 γ 可以包含有限维或无穷维参数, 并设 $p_{\hat\gamma}(x,u)$ 是 $p_\gamma(x,u)$ 的一个估计量. 那么可以用 $p_{\hat\gamma}(x,u)$ 代替 (2.5.7) 式的 $\hat{p}(x,u)$. 参数模型将是首选的例案, 因为 $p_{\hat\gamma}(x,u)$ 容易计算且可用它的渐近分布很好地逼近其真实分布. 诸如部分线性模型和单指标模型等半参数模型也是一个好的选择. 对这些模型的任何一种, 所用方法不要求高维光滑运算.

我们也可以极小化 $\hat{R}_{\rm I}(\beta)$ 得到 β_0 的一个估计量, 记为 $\hat\beta_{\rm I}$, 并称它为基于借补值的极大经验似然估计. 可以推出

$$\hat\beta_{\rm I} = \left(\sum_{i=1}^n \breve{X}_i\breve{X}_i^{\rm T}\right)^{-1}\sum_{i=1}^n \breve{X}_i\breve{Y}_i. \tag{2.5.10}$$

下列定理给出了 $\hat\beta_{\rm I}$ 的渐近正态性.

定理 2.5.4 设条件 (C2.5.1)~(C2.5.10) 成立, 则

$$\sqrt{n}(\hat\beta_{\rm I} - \beta) \xrightarrow{D} N(0,\Gamma),$$

其中 $\Gamma = C^{-1}DC^{-1}$, $C = E\{[X-\mu_2(U)]^{\otimes 2}\}$, $D = E\{p(X,U)^{-1}[X-\mu_2(U)]^{\otimes 2}\varepsilon^2\}$.

为了应用定理 2.5.4 构造 β_0 的置信域, 我们使用 $\hat\Gamma = \hat{C}^{-1}\hat{D}\hat{C}^{-1}$ 作为 Γ 的相合估计, 其中 $\hat{C}$ 和 $\hat{D}$ 分别定义为

$$\hat{C} = \frac{1}{n}\sum_{i=1}^n \breve{X}_i\breve{X}_i^{\rm T}, \quad \hat{D} = \frac{1}{n}\sum_{i=1}^n \hat{p}(X_i,U_i)^{-1}\breve{X}_i\breve{X}_i^{\rm T}(\breve{Y}_i - \hat\beta_{\rm I}^{\rm T}\breve{X}_i)^2.$$

从而可得

$$\hat\Gamma^{-1/2}\sqrt{n}(\hat\beta_{\rm I} - \beta_0) \xrightarrow{D} N(0,I_p), \tag{2.5.11}$$

$$(\hat\beta_{\rm I} - \beta_0)^{\rm T} n\hat\Gamma^{-1}(\hat\beta_{\rm I} - \beta_0) \xrightarrow{D} \chi_p^2. \tag{2.5.12}$$

因此, 使用 (2.5.11) 和 (2.5.12) 可以构造 β_0 的置信区间或置信域.

3. 部分 profile 经验似然

构造 β_0 的某一分量的置信区间也是有兴趣的. 因此, 需要构造部分 profile 经验似然比. 用 e_r 表示第 r 位置的元素为 1 且长度为 p 的单位向量, $r=1,\cdots,p$. 由 (2.5.10) 式, β_0 的第 r 个分量的估计量为 $\hat{\beta}_r=e_r^{\mathrm{T}}\hat{\beta}_{\mathrm{I}}$, $r=1,\cdots,p$. 令

$$\hat{\eta}_{ir,\mathrm{I}}(\beta_r)=e_r^{\mathrm{T}}\hat{C}^{-1}\hat{\eta}_{i,\mathrm{I}}(\hat{\beta}_1,\cdots,\hat{\beta}_{r-1},\beta_r,\hat{\beta}_{r+1},\cdots,\hat{\beta}_p),$$

其中 $\hat{\eta}_{i,\mathrm{I}}(\cdot)$ 在 (2.5.8) 式中定义. 那么, 一个基于借补值的部分 profile 经验似然比函数定义为

$$\hat{R}_{r,\mathrm{I}}(\beta_r)=-2\max\left\{\sum_{i=1}^n\log(np_i)\Big|\,p_i\geqslant 0,\sum_{i=1}^n p_i=1,\sum_{i=1}^n p_i\hat{\eta}_{ir,\mathrm{I}}(\beta_r)=0\right\}.$$

在定理 2.5.3 的条件下, 可以证明 $\hat{R}_{r,\mathrm{I}}(\beta_{0r})$ 的渐近分布是自由度为 1 的 χ^2 分布. 即有下列结果.

定理 2.5.5 设条件 (C2.5.1)~(C2.5.10) 成立, 则对参数真值 β_{0r}, 有 $\hat{R}_{r,\mathrm{I}}(\beta_{0r})\xrightarrow{D}\chi_1^2$.

应用定理 2.5.5, 可定义 β_{0r} 的近似 $1-\alpha$ 置信区间

$$\left\{\beta_r|\,\hat{R}_{r,\mathrm{I}}(\beta_r)\leqslant\chi_1^2(1-\alpha)\right\}.$$

类似地, 通过改进具有完全情况数据的经验似然比, 也可以构造 β_0 的任何分量的置信区间.

2.5.2 缺失响应下基准函数的经验似然

下文假定 u_0 是 $[0,1]$ 的内点, $\sigma_\varepsilon^2(u)=\mathrm{var}(\varepsilon|U=u)$.

1. 估计的经验似然

使用估计量 $\hat{\beta}_{\mathrm{C}}$, 可以构造 $g(u)$ 的估计的经验似然比函数, 即

$$\hat{\mathcal{R}}_{\mathrm{E}}(g(u))=-2\max\left\{\sum_{i=1}^n\log(np_i)\Big|\,p_i\geqslant 0,\sum_{i=1}^n p_i=1,\sum_{i=1}^n p_i\hat{\zeta}_{i,\mathrm{E}}(g(u))=0\right\},$$

其中

$$\hat{\zeta}_{i,\mathrm{E}}(g(u))=\delta_i[Y_i-\hat{\beta}_{\mathrm{C}}^{\mathrm{T}}X_i-g(u)]K\left((U_i-u)/h\right),\qquad(2.5.13)$$

$K(\cdot)$ 和 h 在 (2.5.3) 式中定义.

下列定理陈述了 $\hat{\mathcal{R}}_{\mathrm{E}}(g(u_0))$ 的渐近 χ^2 性.

定理 2.5.6　设条件 (C2.5.1)~(C2.5.5) 成立, 又设方差 $\sigma_\varepsilon^2(u)$ 在点 u_0 处连续, $nh^2/\log n \to \infty$, $nh^5 \to 0$. 如果 $g(u_0)$ 是基准函数的真值, 则 $\hat{\mathcal{R}}_{\rm E}(g(u_0)) \xrightarrow{D} \chi_1^2$.

由定理 2.5.6, 可以定义 $g(u_0)$ 的近似 $1-\alpha$ 置信区间

$$\left\{g(u_0) | \hat{\mathcal{R}}_{\rm E}(g(u_0)) \leqslant \chi_1^2(1-\alpha)\right\}.$$

我们也可以极小化 $\hat{\mathcal{R}}_{\rm E}(g(u))$ 得到 $g(u)$ 的极大经验似然估计量, 记为 $\hat{g}(u)$. 那么 $\hat{g}(u)$ 可以表示为

$$\hat{g}(u) = \hat{\mu}_1(u) - \hat{\beta}_{\rm C}^{\rm T}\hat{\mu}_2(u), \tag{2.5.14}$$

其中 $\hat{\mu}_1(u)$ 和 $\hat{\mu}_2(u)$ 在 (2.5.2) 式中定义.

定理 2.5.7　设条件 (C2.5.1)~(C2.5.6) 成立, 方差 $\sigma_\varepsilon^2(u)$ 在点 u_0 处连续, 则

$$\sqrt{nh}[\hat{g}(u_0) - g(u_0)] - b(u_0)[q(u_0)f(u_0)]^{-1} \xrightarrow{D} N(0, \gamma^2(u_0)),$$

其中 $\gamma^2(u_0) = v^2(u_0)[q(u_0)f(u_0)]^{-2}$, $b(u_0)$ 和 $v^2(u_0)$ 分别定义为

$$\begin{aligned} b(u_0) = {} & h_0^{5/2}[g'(u_0)q'(u_0)f(u_0) + g'(u_0)q(u_0)f'(u_0) \\ & +(1/2)g''(u_0)q(u_0)f(u_0)]\int_{-1}^{1} u^2K(u){\rm d}u, \end{aligned} \tag{2.5.15}$$

$$v^2(u_0) = \sigma_\varepsilon^2(u_0)q(u_0)f(u_0)\int_{-1}^{1} K^2(u){\rm d}u, \tag{2.5.16}$$

h_0 是由条件 (C2.5.5) 给定的常数. 进一步, 如果用 $nh^2/\log n \to \infty$ 和 $nh^5 \to 0$ 代替条件 (C2.5.5), 则偏差 $b(u_0)$ 消失, 即有

$$\sqrt{nh}[\hat{g}(u_0) - g(u_0)] \xrightarrow{D} N(0, \gamma^2(u_0)).$$

下面考虑 $g(u_0)$ 的逐点置信区间的构造. 由定理 2.5.7, 可以构造 $b(u_0)$ 和 $\gamma^2(u_0)$ 的估计量. 注意到

$$\sqrt{nh}E\{\delta[g(U) - g(u_0)]K_h(U-u_0)\} = b(u_0) + o(1).$$

因此, $b(u_0)$ 的相合估计量定义为

$$\hat{b}(u_0) = (h/n)^{1/2}\sum_{i=1}^{n}\delta_i[\hat{g}(U_i) - \hat{g}(u_0)]K_h(U_i - u_0). \tag{2.5.17}$$

下面定义 $\gamma^2(u_0)$ 的估计量. 因为 $\gamma^2(u_0)$ 依赖于 $f(u_0)$、$q(u_0)$ 和 $v^2(u_0)$, 所以我们需要给出它们的估计. 使用核函数 $K(\cdot)$ 和带宽 h, 可以定义 $f(u_0)$, $q(u_0)$ 和 $v(u_0)$ 的估计量分别为

$$\hat{f}(u_0) = \frac{1}{n}\sum_{i=1}^{n}K_h(U_i - u_0),$$

$$\hat{q}(u_0) = \frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u_0)\delta_i/\hat{f}(u_0),$$

$$\hat{v}^2(u_0) = \frac{1}{nh}\sum_{i=1}^{n} \hat{\zeta}_{i,\mathrm{E}}^2(\hat{g}(u_0)), \tag{2.5.18}$$

其中 $\hat{\zeta}_{i,\mathrm{E}}(\hat{g}(u_0))$ 在 (2.5.13) 式中定义. 因此, 用 $\hat{f}(u_0)$, $\hat{q}(u_0)$ 和 $\hat{v}^2(u_0)$ 分别代替 $\gamma^2(u_0)$ 中的 $f(u_0)$, $q(u_0)$ 和 $v^2(u_0)$, 可以得到 $\gamma^2(u_0)$ 的一个估计量, 并记为 $\hat{\gamma}^2(u_0)$. 可以证明, $\hat{f}(u_0)$, $\hat{q}(u_0)$ 和 $\hat{v}^2(u_0)$ 分别是 $f(u_0)$, $q(u_0)$ 和 $v^2(u_0)$ 的相合估计. 因此, $\hat{\gamma}^2(u_0)$ 是 $\gamma^2(u_0)$ 的相合估计. 此外, 引理 2.5.5 已证明 $\hat{b}(u_0)$ 是 $b(u_0)$ 的相合估计. 因此, 由定理 2.5.7 可得

$$\hat{\gamma}^{-1}(u_0)\left\{\sqrt{nh}[\hat{g}(u_0) - g(u_0)] - \hat{b}(u_0)[\hat{q}(u_0)\hat{f}(u_0)]^{-1}\right\} \xrightarrow{D} N(0,1). \tag{2.5.19}$$

使用 (2.5.19) 式可以得到 $g(u_0)$ 的近似 $1-\alpha$ 置信区间, 即

$$\hat{g}(u_0) - (nh)^{-1/2}\hat{b}(u_0)[\hat{q}(u_0)\hat{f}(u_0)]^{-1} \pm z_{1-\alpha/2}(nh)^{-1/2}\hat{\gamma}(u_0),$$

其中 $z_{1-\alpha/2}$ 是标准正态分布的 $1-\alpha/2$ 分位数.

进一步, 如果选取带宽为 $h = o(n^{-1/5})$, 那么 $\hat{g}(u_0)$ 的偏差可忽略, 且可定义 $g(u_0)$ 的近似 $1-\alpha$ 置信区间

$$\hat{g}(u_0) \pm z_{1-\alpha/2}(nh)^{-1/2}\hat{\gamma}(u_0).$$

2. *残差调整的经验似然*

我们可以调整加权的残差 $\hat{\eta}_{i,\mathrm{E}}(g(u))$ 得到一个残差调整的经验似然比函数. 为此, 引入辅助随机变量

$$\hat{\zeta}_{i,\mathrm{R}}(g(u)) = \delta_i\left[Y_i - \hat{\beta}_{\mathrm{C}}^{\mathrm{T}}X_i - g(u) - [\hat{g}(U_i) - \hat{g}(u)]\right]K\left((U_i - u)/h\right),$$

其中 $K(\cdot)$ 和 h 在 (2.5.3) 式中定义. 因此, $g(u)$ 的残差调整的经验似然比函数定义为

$$\hat{\mathcal{R}}_{\mathrm{R}}(g(u)) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\middle| p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\hat{\zeta}_{i,\mathrm{R}}(g(u)) = 0\right\}.$$

通过 $\hat{\zeta}_{i,\mathrm{R}}(g(u))$ 中的残差调整, 我们不仅减少了偏差, 而且避免了对基准函数的欠光滑. 也就是说, 不需要欠光滑即可得到 $\hat{\mathcal{R}}_{\mathrm{R}}(g(u_0))$ 的渐近 χ^2 性. 下列定理陈述了这个结果.

定理 2.5.8 设条件 (C2.5.1)~(C2.5.6) 成立, 又设方差 $\sigma_\varepsilon^2(u)$ 在点 u_0 处连续, 核 $K(\cdot)$ 在 $[0,1]$ 上二次连续可微. 如果 $g(u_0)$ 是基准函数真值, 则 $\hat{\mathcal{R}}_{\rm R}(g(u_0)) \xrightarrow{D} \chi_1^2$.

由定理 2.5.8, 可以定义 $g(u_0)$ 的近似 $1-\alpha$ 置信区间

$$\left\{\tilde{g}(u_0)|\, \hat{\mathcal{R}}_{\rm R}(\tilde{g}(u_0)) \leqslant \chi_1^2(1-\alpha)\right\}. \tag{2.5.20}$$

利用 (2.5.20) 可以构造函数 $g(u)$ 的逐点置信带.

3. 同时置信带

如何将 $g(u)$ 的逐点置信带调整为同时置信带是一个有趣的问题. 为解决这个问题, 下面介绍一个简单方法. 首先考虑基于残差调整的经验似然. 用 $[\hat{L}_\alpha(u_0), \hat{U}_\alpha(u_0)]$ 表示由 (2.5.20) 式定义的逐点置信区间. 设 $[a,b]$ 是 $[0,1]$ 的子区间, M 为使得 $M>(b-a)/h$ 的最小整数. 将 $[a,b]$ 分成 $M+1$ 个等距的格子点 $a=\xi_1<\cdots<\xi_{M+1}=b$. 可以得到 $g(u)$ 的一组近似 $1-\alpha$ 同时置信区间 $[\hat{l}_\alpha(\xi_k), \hat{u}_\alpha(\xi_k)]$, 使得

$$\lim_{n\to\infty} P\left\{\hat{l}_\alpha(\xi_k) \leqslant g(\xi_k) \leqslant \hat{u}_\alpha(\xi_k), \forall\, k=1,\cdots,M+1\right\} \geqslant 1-\alpha.$$

特别地, 如果使用 Bonferroni 调整, $[\hat{l}_\alpha(\xi_k), \hat{u}_\alpha(\xi_k)]$ 可由 $[\hat{L}_{\alpha/(M+1)}(\xi_k), \hat{U}_{\alpha/(M+1)}(\xi_k)]$ 给出. 设 $g^{(I)}(u)$ 是 $g(\xi_k)$ 和 $g(\xi_{k+1})$ 的线性组合, 即对 $u\in[\xi_k,\xi_{k+1}]$, 使得

$$g^{(I)}(u) = M\left(\frac{\xi_{k+1}-u}{b-a}\right)g(\xi_k) + M\left(\frac{u-\xi_k}{b-a}\right)g(\xi_{k+1}).$$

类似地, 设 $\hat{l}_\alpha^{(I)}(u)$ 和 $\hat{u}_\alpha^{(I)}(u)$ 分别是 $\hat{l}_\alpha(\xi_k)$ 与 $\hat{l}_\alpha(\xi_{k+1})$ 和 $\hat{u}_\alpha(\xi_k)$ 与 $\hat{u}_\alpha(\xi_{k+1})$ 的线性组合. 那么, $\left[\hat{l}_\alpha^{(I)}(u), \hat{u}_\alpha^{(I)}(u)\right]$ 是 $g^{(I)}(u)$ 的近似 $1-\alpha$ 置信带, 且有

$$\lim_{n\to\infty} P\left\{\hat{l}_\alpha^{(I)}(u) \leqslant g^{(I)}(u) \leqslant \hat{u}_\alpha^{(I)}(u), \forall\, t\in[a,b]\right\} \geqslant 1-\alpha.$$

由条件 (C2.5.2) 可知 $g(u)$ 是 $[a,b]$ 上的有界函数, 并记 $c=\sup\limits_{t\in[a,b]}|g''(u)|$. 应用 Taylor 公式可得: 对所有 $u\in[\xi_k,\xi_{k+1}]$,

$$|g(u)-g^{(I)}(u)| \leqslant c(\xi_{k+1}-u)(u-\xi_k)/2.$$

调整 $g^{(I)}(u)$ 的带, 可得到 $g(u)$ 在 $[a,b]$ 上的近似 $1-\alpha$ 置信带: 对所有 $u\in[\xi_k,\xi_{k+1}]$, $k=1,\cdots,M$,

$$\left[\hat{l}_\alpha^{(I)}(u) - c(\xi_{k+1}-u)(u-\xi_k)/2,\ \hat{u}_\alpha^{(I)}(u) + c(\xi_{k+1}-u)(u-\xi_k)/2\right]. \tag{2.5.21}$$

类似地, 我们也可以对估计的经验似然和正态逼近构造 $g(u)$ 的同时置信带. 这里不再一一赘述.

注 2.5.2 对 (2.5.21) 式定义的置信带, M 的选择是一个重要的问题. Hall 和 Titterington (1988) 建议取 M 为满足 $M > (b-a)/h$ 的最小整数. 他们针对单纯的固定设计非参数曲线给出了置信带的宽度收敛于 0 的最优速度, 且证明了如果正确选择格子点数可达到他们的最优速度.

2.5.3 缺失响应下响应均值的估计

本节使用经验似然提出对响应均值 θ 作整体推断的方法. 我们使用纠偏技术和加权半参数回归借补来直接构造 θ 的经验对数似然比统计量, 使得该统计量是渐近 χ^2 的. 因此, 本节的方法不同于 Wang 等 (2004) 的方法. 因为他们使用的方法需乘一个调整因子来得到经验似然比, 当有效地估计调整因子遇到困难时就会带来麻烦.

1. *加权校正的经验似然*

我们现在提出一个加权校正的经验似然方法对 θ 进行整体推断. 设

$$Y_i^* = \frac{\delta_i Y_i}{p(X_i,U_i)} + \left(1 - \frac{\delta_i}{p(X_i,U_i)}\right)\left[\beta_0^{\mathrm{T}} X_i + g(U_i)\right],$$

在 MAR 下, 如果 θ 是参数真值, 那么 $E(Y_i^*) = \theta$. 因此, 可定义经验似然比函数 $l^*(\theta)$. 然而, β_0, $p(\cdot,\cdot)$ 和 $g(\cdot)$ 通常是未知的, 我们需要用它们的估计量分别代替这些未知量. 由 (2.5.4), (2.5.7) 和 (2.5.14) 式, 用 $\hat{\beta}_{\mathrm{C}}$, $\hat{p}(X_i,U_i)$ 和 $\hat{g}(U_i)$ 代替 Y_i^* 中的 β_0, $p(\cdot,\cdot)$ 和 $g(\cdot)$ 可得到 Y_i^* 的一个估计量 $\hat{Y}_i$. 即

$$\hat{Y}_i = \frac{\delta_i Y_i}{\hat{p}(X_i,U_i)} + \left(1 - \frac{\delta_i}{\hat{p}(X_i,U_i)}\right)\left[X_i^{\mathrm{T}}\hat{\beta}_{\mathrm{C}} + \hat{g}(U_i)\right]. \tag{2.5.22}$$

因此, θ 的加权校正的经验对数似然比函数定义为

$$\hat{l}(\theta) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\,\middle|\, p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\hat{Y}_i = \theta\right\}.$$

下列定理 2.5.9 表明 $\hat{l}(\theta)$ 的渐近分布是自由度为 1 的 χ^2 分布.

定理 2.5.9 设条件 (C2.5.1)~(C2.5.10) 成立. 如果 θ 是参数真值, 则 $\hat{l}(\theta) \xrightarrow{D} \chi_1^2$.

注 2.5.3 如果去掉 (2.5.22) 式中的 $\hat{p}(X_i,U_i)$, 那么相应的经验对数似然比将渐近于比例 χ^2 变量. 因为使用了加权校正, 所以 $\hat{l}(\theta)$ 具有标准 χ^2 分布.

使用定理 2.5.9, 可定义 θ 的近似 $1-\alpha$ 置信区间

$$\left\{\tilde{\theta} \mid \hat{l}(\tilde{\theta}) \leqslant \chi_1^2(1-\alpha)\right\}.$$

定理 2.5.9 也可以用来检验假设 $H_0: \theta = \theta_0$. 如果 $\hat{l}(\theta_0) > \chi_1^2(1-\alpha)$, 那么以水平 α 拒绝 H_0.

2. 正态逼近

Xue 等 (2011) 提出了 θ 的加权借补的估计量, 即

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i, \tag{2.5.23}$$

其中 $\hat{Y}_i$ 在 (2.5.22) 式中定义.

注 2.5.4　估计量 $\hat{\theta}$ 不同于 Wang 等 (2004) 的估计量 $\hat{\theta}_P$, 因为 $\hat{\theta}$ 中的估计量 $\hat{p}(x,u)$ 不同于他们的估计量. 此外, 估计量 $\hat{\theta}$ 也不同于 Liang 等 (2007) 的估计量 $\hat{\theta}_{n,\mathrm{wei}}$, 因为他们在 $\hat{\theta}_{n,\mathrm{wei}}$ 中使用选择概率估计 $\hat{p}(u)$, 而不是使用估计量 $\hat{p}(x,u)$.

我们也可以极小化 $\hat{l}(\theta)$ 得到 θ 的具有完全情况数据的极大经验似然估计 $\hat{\theta}_{\mathrm{ME}}$. 可以证明 $\hat{\theta}_{\mathrm{ME}} = \hat{\theta}$. 下列定理中给出了 $\hat{\theta}$ 的渐近分布.

定理 2.5.10　设条件 (C2.5.1)~(C2.5.10) 成立, 则

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{D} N(0,V),$$

其中

$$V = E\left[\frac{\sigma^2(X,U)}{p(X,U)}\right] + \mathrm{var}(X^{\mathrm{T}}\beta + g(u)),$$

$\sigma^2(x,u) = \mathrm{var}(\varepsilon|X=x, U=u)$.

定义 V 的估计量

$$\hat{V} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - \hat{\theta})^2.$$

可以证明 $\hat{V}$ 是 V 的相合估计. 由定理 2.5.10, 可以定义 θ 的正态逼近 $1-\alpha$ 置信区间 $\hat{\theta} \pm z_{1-\alpha/2}(\hat{V}/n)^{1/2}$.

2.5.4　缺失协变量下回归系数的估计

本节假定协变量 X 随机缺失 (MAR). MAR 假定蕴含给定 Y 和 U 下 δ 与 X 条件独立, 即

$$P(\delta=1|X,U,Y) = P(\delta=1|Y,U) \equiv \pi(Y,U). \tag{2.5.24}$$

MAR 是对缺失数据统计分析的普通假定, 它在许多情况下是合理的. 参见文献 (Little and Rubin, 1987, Chapter 1). 本节用两种不同方法定义 β_0 和 $g(\cdot)$ 的估计量, 一是逆概率加权方法, 另一个是逆概率加权借补方法. 我们得到了 β_0 的估计量的渐近正态性, 并给出了 $g(\cdot)$ 的估计量的最优收敛速度.

1. 逆概率加权方法

如果 $\pi(y,u)$ 和 $g(u)$ 是已知的, 那么简单计算可得

$$E\left\{\frac{\delta}{\pi(Y,U)}[Y-\beta_0^{\mathrm{T}}X-g(U)]^2\right\}=E\left\{[Y-\beta_0^{\mathrm{T}}X-g(U)]^2\right\}.$$

因此, 我们可以极小化目标函数

$$Q(\beta)=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(Y_i,U_i)}[Y_i-\beta^{\mathrm{T}}X_i-g(U_i)]^2, \tag{2.5.25}$$

得到 β_0 的估计量. 然而, $\pi(y,u)$ 和 $g(u)$ 是未知的, 我们需要用它们的估计量来代替. 可以用任何一种光滑方法估计 $\pi(y,u)$ 和 $g(u)$, 这里采用核估计法. 选定核函数 $K(\cdot)$ 和 $K^*(\cdot)$ 以及收敛于 0 的带宽 $h_1=h_{1n}$ 和 $h_2=h_{2n}$, 记 $K_{h_1}(u)=h_1^{-1}K(u/h_1)$, $K_{h_2}^*(y,u)=h_2^{-2}K^*(y/h_2,u/h_2)$, $\xi=X\delta/\pi(Y,U)$. 注意到

$$g(u)=E(Y-\beta_0^{\mathrm{T}}X|U=u)=E(Y-\beta_0^{\mathrm{T}}\xi|U=u).$$

于是, 对固定的 β_0, 可定义 $\pi(y,u)$ 和 $g(u)$ 的核估计分别为

$$\hat{\pi}(y,u)=\sum_{i=1}^{n}W_{ni}^*(y,u)\delta_i,$$

$$\hat{g}(u;\beta_0)=\sum_{i=1}^{n}W_{ni}(u)[Y_i-\beta_0^{\mathrm{T}}X_i\delta_i/\hat{\pi}(Y_i,U_i)],$$

其中

$$W_{ni}(u)=\frac{K_{h_1}(U_i-u)}{\sum\limits_{j=1}^{n}K_{h_1}(U_j-u)}.$$

$$W_{ni}^*(y,u)=\frac{K_{h_2}^*(Y_i-y,U_i-u)}{\sum\limits_{j=1}^{n}K_{h_2}^*(Y_j-y,U_j-u)}, \tag{2.5.26}$$

将 $\hat{\pi}(y,u)$ 和 $\hat{g}(u;\beta)$ 代入 (2.5.25) 式可得

$$\hat{Q}(\beta)=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(Y_i,U_i)}[Y_i-\beta^{\mathrm{T}}X_i-\hat{g}(U_i;\beta)]^2.$$

极小化 $\hat{Q}(\beta)$ 即可得到 β_0 的估计量, 称之为逆概率加权估计, 并记为 $\hat{\beta}_{\mathrm{IW}}$. 可以证明, $\hat{\beta}_{\mathrm{IW}}$ 是加权估计方程

$$\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(Y_i,U_i)}\widetilde{X}_i[\widetilde{Y}_i-\widetilde{X}_i^{\mathrm{T}}\beta]=0 \tag{2.5.27}$$

的解, 其中 $\widetilde{Y}_i = Y_i - \hat{g}_1(U_i)$, $\widetilde{X}_i = X_i - \hat{g}_2(U_i)$, $1 \leqslant i \leqslant n$,

$$\hat{g}_1(u) = \sum_{i=1}^{n} W_{ni}(u) Y_i, \quad \hat{g}_2(u) = \sum_{i=1}^{n} \frac{X_i \delta_i}{\hat{\pi}(Y_i, U_i)} W_{ni}(u).$$

这里 $\hat{g}_1(u)$ 和 $\hat{g}_2(u)$ 分别为 $g_1(u) = E(Y|U=u)$ 和 $g_2(u) = E(X|U=u)$ 的估计量. 解方程 (2.5.27) 可得

$$\hat{\beta}_{\mathrm{IW}} = \hat{\Sigma}_0^{-1} \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(Y_i, U_i)} \widetilde{X}_i \widetilde{Y}_i, \tag{2.5.28}$$

其中

$$\hat{\Sigma}_0 = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(Y_i, U_i)} \widetilde{X}_i \widetilde{X}_i^{\mathrm{T}}.$$

利用 $\hat{\beta}_{\mathrm{IW}}$, 可定义 $g(u)$ 的最终估计量 $\hat{g}_{\mathrm{IW}}(u) = \hat{g}(u; \hat{\beta}_{\mathrm{IW}})$.

为得到 $\hat{\beta}_{\mathrm{IW}}$ 的渐近正态性, 需给出下列条件.

(D2.5.1) 选择概率 $\pi(y,u)$ 几乎处处有 $k(>2)$ 阶有界偏导数, 且 $\inf\limits_{y,u} \pi(y,u) > 0$.

(D2.5.2) 随机变量 (Y,U) 的密度函数 $f(y,u)$ 几乎处处有 $k(>2)$ 阶有界偏导数. $f(y,u)$ 和 U 的密度函数 $f(u)$ 有界远离 0.

(D2.5.3) 函数 $g(u)$, $g_1(u)$ 和 $g_{2s}(u)$ 在 $(0,1)$ 上具有有界连续的 2 阶导数, 其中 $g_{2s}(u)$ 为 $g_2(u)$ 的第 s 个分量.

(D2.5.4) $\sup\limits_{x,u} E(\varepsilon^4|X=x, U=u) < \infty$, $\sup\limits_{u} E(||X||^4|U=u) < \infty$.

(D2.5.5) 存在正的常数 C_1, C_2 和 ρ 使得 $C_1 I(|u| \leqslant \rho) \leqslant K(u) \leqslant C_2 I(|u| \leqslant \rho)$.

(D2.5.6) 核 $K^*(u)$ 是 $k(>2)$ 阶有界函数, 且具有有界支撑.

(D2.5.7) 带宽满足: $nh_1^2 \to \infty$, $nh_1^4 \to 0$, $nh_2^4 \to \infty$, $nh_1^2 h_2 \to \infty$, $nh_2^{2k+1} \to 0$, $h_1/h_2 \to 0$, $h_2^k/h_1 \to 0$, $k > 2$.

下列定理给出了 $\hat{\beta}_{\mathrm{IW}}$ 的渐近正态性.

定理 2.5.11　设条件(D2.5.1)~(D2.5.7) 成立, 函数 $M(y,u)$ 几乎处处有 $k(>2)$ 阶有界偏导数, Σ_0 是正定矩阵, 则

$$\sqrt{n}(\hat{\beta}_{\mathrm{IW}} - \beta_0) \xrightarrow{D} N(0, \Sigma_0^{-1} \Omega \Sigma_0^{-1}),$$

其中 $\Sigma_0 = E\{[X - g_2(U)]^{\otimes 2}\}$, $\Omega = E\{[X - g_2(U)]^{\otimes 2} \varepsilon^2\} + E\{[1 - \pi^{-1}(Y,U)] M(Y,U)^{\otimes 2}\}$, $M(y,u) = E[(X - g_2(U))\varepsilon | Y=y, U=u]$.

为应用定理 2.5.11 构造 β_0 的置信域, 需要估计 Σ_0 和 Ω. Σ_0 的估计量由 (2.5.28) 式定义. 由嵌入 (plug-in) 方法可得 Ω 的估计量为

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\delta_i}{\hat{\pi}(Y_i, U_i)} \widetilde{X}_i \widetilde{X}_i^{\mathrm{T}} \hat{\varepsilon}_i^2 + \left[1 - \frac{1}{\hat{\pi}(Y_i, U_i)}\right] \hat{M}(Y_i, U_i)^{\otimes 2} \right\},$$

其中

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_{\rm IW}^{\rm T} X_i \delta_i / \hat{\pi}(Y_i, U_i) - \hat{g}_{\rm IW}(U_i),$$

$$\hat{M}(y,u) = \sum_{i=1}^{n} W_{ni}^*(y,u) \widetilde{X}_i \hat{\varepsilon}_i,$$

$W_{ni}^*(y,u)$ 由 (2.5.26) 式定义.

下列定理给出了 $\hat{g}_{\rm IW}(u)$ 的收敛速度.

定理 2.5.12 在定理 2.5.11 的条件下, 如果 $h_1 = O(n^{-1/3})$ 和 $h_2 = O(n^{-1/6})$, 则

$$\hat{g}_{\rm IW}(u) - g(u) = O_P(n^{-1/3}).$$

估计量 $\hat{g}_{\rm IW}(u)$ 的收敛速度与 Stone(1980) 得到的最优收敛速度相同.

2. *逆概率加权借补方法*

逆概率加权方法仅使用观察到的数据来构造 β_0 的估计量, 丢掉了部分缺失数据, 从而会降低估计精度. 为解决这个问题, 一个通常的方法是利用预测值填补缺失值. 记

$$\varphi(Y_i, U_i; \beta) = E[\zeta_i(\beta) | Y_i, U_i], \quad \zeta_i(\beta) = \widetilde{X}_i(\widetilde{Y}_i - \widetilde{X}_i^{\rm T}\beta),$$

其中 $\widetilde{X}_i$ 和 $\widetilde{Y}_i (1 \leqslant i \leqslant n)$ 在 (2.5.27) 式中定义. 按照 Robins 等 (1994) 的思想, 我们使用 $\varphi(Y_i, U_i; \beta)$ 填补 $\zeta_i(\beta)$. 因为 $\varphi(Y_i, U_i; \beta)$ 是未知的, 我们需要估计它. 对给定的 β, 定义 $\varphi(y, u; \beta)$ 的核估计

$$\hat{\varphi}(y, u; \beta) = \sum_{i=1}^{n} W_{ni}^*(y,u) \zeta_i(\beta),$$

其中 $W_{ni}^*(y,u)$ 由 (2.5.26) 式定义. 因此, 构造估计方程

$$\sum_{i=1}^{n} \left\{ \frac{\delta_i}{\hat{\pi}(Y_i, U_i)} \zeta_i(\beta) + \left[1 - \frac{\delta_i}{\hat{\pi}(Y_i, U_i)}\right] \hat{\varphi}(Y_i, U_i; \beta) \right\} = 0. \tag{2.5.29}$$

解上述方程可得到 β_0 的估计量, 记为 $\hat{\beta}_{\rm IP}$, 并称之为 β_0 的逆概率加权借补估计. 可以证明下列结果.

定理 2.5.13 在定理 2.5.11 的条件下, 又设 $E(X|Y=y, U=u)$ 和 $E(XX^{\rm T}|Y=y, U=u)$ 具有有界的 1 阶导数, 则估计量 $\hat{\beta}_{\rm IP}$ 与 $\hat{\beta}_{\rm IW}$ 有相同的渐近正态分布.

3. *经验似然*

为构造任何参数 β 的经验似然比, 需引入辅助随机变量

$$\widetilde{\eta}_i(\beta) = \frac{\delta_i}{\hat{\pi}(Y_i, U_i)} \widetilde{X}_i(\widetilde{Y}_i - \widetilde{X}_i^{\rm T}\beta),$$

其中 $\widetilde{X}_i$ 和 $\widetilde{Y}_i(1 \leqslant i \leqslant n)$ 在 (2.5.27) 式中定义. 因此, β 的逆概率加权经验似然比函数定义为

$$\mathcal{R}_{\rm IW}(\beta) = -2\max\left\{\sum_{i=1}^n \log(np_i)\Big| p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\widetilde{\eta}_i(\beta) = 0\right\}.$$

我们也可以引入辅助随机变量

$$\hat{\eta}_i(\beta) = \frac{\delta_i}{\hat{\pi}(Y_i, U_i)}\zeta_i(\beta) + \left[1 - \frac{\delta_i}{\hat{\pi}(Y_i, U_i)}\right]\hat{\varphi}(Y_i, U_i; \beta).$$

从而可构造 β_0 的逆概率加权借补经验似然比函数为

$$\mathcal{R}_{\rm IP}(\beta) = -2\max\left\{\sum_{i=1}^n \log(np_i)\Big| p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\hat{\eta}_i(\beta) = 0\right\}.$$

可以证明下列结果.

定理 2.5.14　在定理 2.5.11 的条件下, 有 $\mathcal{R}_{\rm IW}(\beta_0) \xrightarrow{D} \chi^2_{p-1}$, $\mathcal{R}_{\rm IP}(\beta_0) \xrightarrow{D} \chi^2_{p-1}$.

利用上述结果可以构造 β_0 的大样本置信区间, 并可以进行假设检验.

2.5.5　模拟研究

1. *一维情形*

利用模型 (2.1.1) 产生模拟数据, 其中 $\beta_0 = 1.5$, $g(u) = [4(u-0.5)^2 + 0.5]I(0 < u < 1)$. 此外, 变量 X 和 U 分别服从正态分布 $N(1,1)$ 和均匀分布 $U(0,1)$, ε 服从标准正态分布 $N(0,1)$. 核函数取为 $K(u) = (15/16)(1-u^2)^2 I(|u| \leqslant 1)$ 和 $\widetilde{K}(x,u) = K(x)K(u)$. 使用交叉验证 (CV) 方法选择最优带宽 $h_{\rm opt}$ 和 $b_{\rm opt}$, 且取 $M_n = \ln n + 1$. 不难验证它们满足条件 (C2.5.9) 和 (C2.5.10).

基于下列 4 个选择概率函数 $p(x,u)$, 分别产生样本量为 $n = 100, 150, 250$ 的 5000 个随机样本:

情况 1. 如果 $(x-1)^2 + (u-0.5)^2 < 1$, 则 $p_1(x,u) = 0.8 + 0.2[(x-1)^2 + (u-0.5)^2]$, 否则 $p_1(x,u) = 0.96$.

情况 2. 如果 $(x-1)^2 + (u-0.5)^2 < 1$, 则 $p_2(x,u) = 0.7 + 0.2[(x-1)^2 + (u-0.5)^2]$, 否则 $p_2(x,u) = 0.72$.

情况 3. 如果 $(x-1)^2 + (u-0.5)^2 < 1$, 则 $p_3(x,u) = 0.6 + 0.2[(x-1)^2 + (u-0.5)^2]$, 否则 $p_3(x,u) = 0.62$.

情况 4. 对所有 x 和 u, $p_4(x,u) = 1$.

对于上面四种情况, 平均响应速率为 $E[p_1(X,U)] \approx 0.90$, $E[p_2(X,U)] \approx 0.75$, $E[p_3(X,U)] \approx 0.65$, $E[p_4(X,U)] = 1$. 情况 4 不是缺失数据, 它是用来作为整体标准. 对下列三种情形进行模拟研究.

(a) 我们使用三种方法来构造 β_0 的置信区间, 即, 完全情况数据的经验似然 (CEL), 借补值的经验似然 (IEL) 和正态逼近 (NA). 下文用 NA($\hat{\beta}_C$) 和 NA($\hat{\beta}_I$) 分别表示 $\hat{\beta}_C$ 和 $\hat{\beta}_I$ 对应的正态逼近置信区间. 用 5000 次模拟运行来计算 β_0 的近似 0.95 置信区间的平均长度和经验覆盖概率, 其结果列在表 2.5.1 中.

表 2.5.1 当维数 $p=1$ 时, 在不同的选择概率函数 $p(x,u)$ 和样本量 n 下, β_0 的近似 0.95 置信区间的平均长度和经验覆盖概率

$p(x,u)$	n	平均长度				覆盖概率			
		IEL	CEL	NA($\hat{\beta}_I$)	NA($\hat{\beta}_C$)	IEL	CEL	NA($\hat{\beta}_I$)	NA($\hat{\beta}_C$)
$p_1(x,u)$	100	0.4148	0.4042	0.4076	0.3982	0.9344	0.9320	0.9318	0.9292
	150	0.3373	0.3296	0.3326	0.3256	0.9468	0.9418	0.9438	0.9376
	250	0.2597	0.2554	0.2572	0.2531	0.9474	0.9446	0.9462	0.9426
$p_2(x,u)$	100	0.4468	0.4593	0.4317	0.4521	0.9234	0.9312	0.9126	0.9288
	150	0.3659	0.3746	0.3563	0.3695	0.9298	0.9384	0.9212	0.9356
	250	0.2838	0.2904	0.2784	0.2874	0.9322	0.9402	0.9282	0.9384
$p_3(x,u)$	100	0.4755	0.4936	0.4548	0.4853	0.9124	0.9278	0.9010	0.9248
	150	0.3907	0.4035	0.3775	0.3976	0.9258	0.9382	0.9164	0.9340
	250	0.3038	0.3125	0.2964	0.3090	0.9330	0.9400	0.9272	0.9376
$p_4(x,u)$	100	0.3947	0.3947	0.3892	0.3890	0.9322	0.9332	0.9294	0.9292
	150	0.3221	0.3218	0.3183	0.3180	0.9422	0.9416	0.9394	0.9392
	250	0.2494	0.2492	0.2472	0.2471	0.9428	0.9428	0.9404	0.9408

从表 2.5.1 可以得到如下结果. (i) 对情况 1, IEL 比其他两种方法有较大的覆盖概率, 但有较长的区间长度. 对应三种方法的区间长度比情况 4 仅稍微长, 因此所提出的方法执行效果好. 对情况 2 和 3, IEL 比 CEL 的区间长度短, 但覆盖概率大. 对所有四种情况, IEL 和 CEL 分别比 NA($\hat{\beta}_C$) 和 NA($\hat{\beta}_I$) 有大的覆盖概率, 但有较长的区间长度. (ii) 当 n 增大时, 四种情况对应的置信区间的经验覆盖概率增大而区间长度减小. (iii) 覆盖概率和区间长度依赖于选择概率函数. 情况 1 与情况 2 和 3 相比, 所有方法的执行效果都很好. 这是由于情况 1 的缺失速度比情况 2 和 3 的缺失速度低. 一般来说, 当缺失速度增大时覆盖概率减小而区间长度增大.

(b) 用三个方法构造 $g(u)$ 的置信带: 估计的经验似然 (EEL), 残差调整的经验似然 (RAEL) 和 NA. 对于 NA, $g(u)$ 的估计量使用 (2.5.14) 式定义的 $\hat{g}(u)$. 基于上述四个选择概率函数以及 $n=100$, 通过 400 次模拟运行来计算 $g(u)$ 的近似 0.95 置信带. 模拟结果展示在图 2.5.1 中.

从图 2.5.1 可以看出 RAEL 比 EEL 和 NA 的执行效果明显好, 因为 RAEL 的

置信带窄.

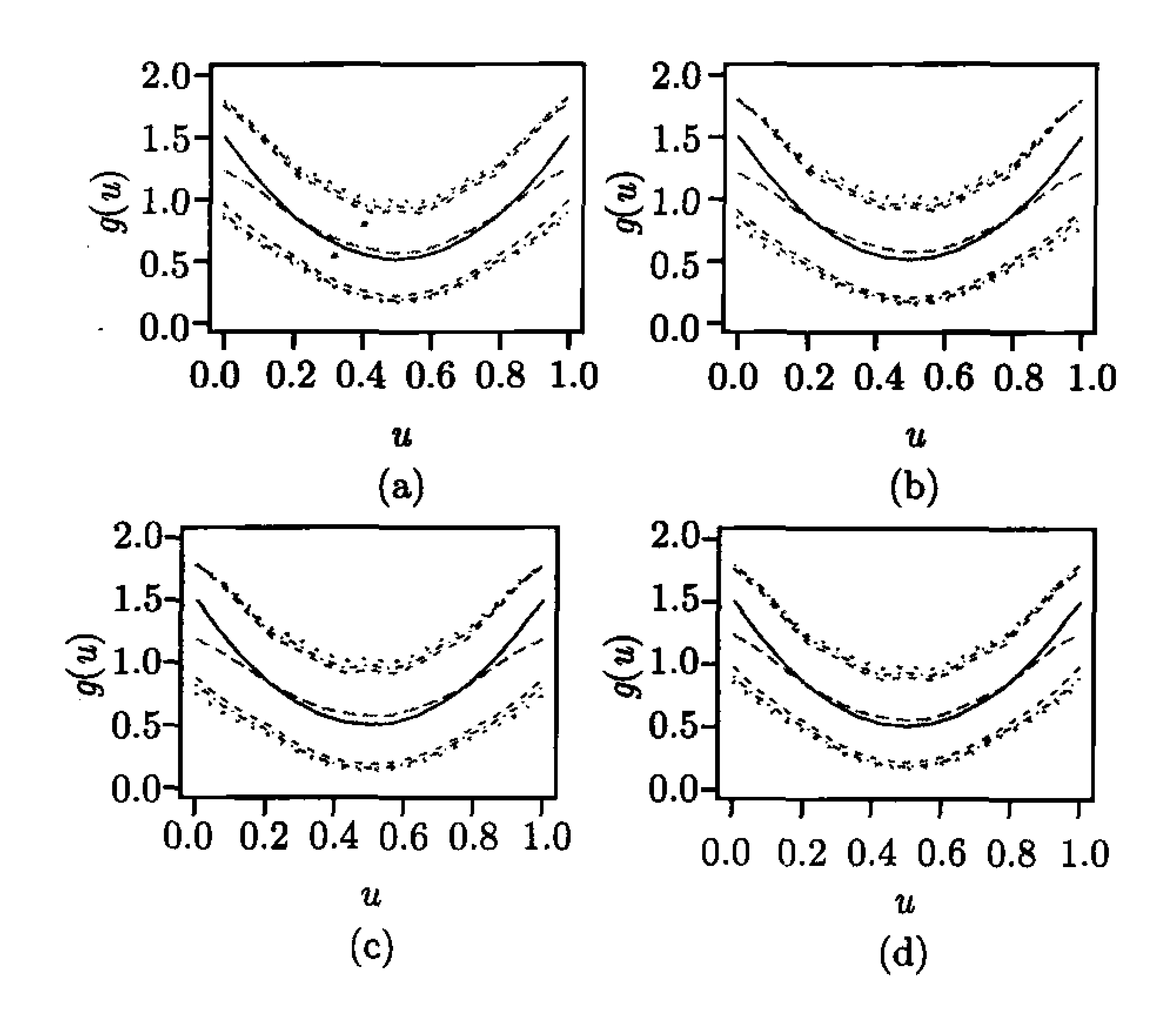


图 2.5.1　基于 RAEL(虚线)、EEL(点线) 和 NA(点虚线) 以及 $n=100$ 和 $p=1$, $g(u)$ 的近似 0.95 置信带. 实线表示 $g(u)$ 的真实曲线, 长虚线表示 $g(u)$ 的估计的曲线. (a)~(d) 分别对应情况 1~4

(c) 为得到 θ 的置信区间, 我们考虑三种方法: 基于 $\hat{l}(\theta)$ 的加权校正的经验似然 (WCEL), Wang 等 (2004) 提出的调整的经验似然 (AEL) 和 NA. 对于 NA, 使用 (2.5.23) 式定义的 $\hat{\theta}$ 作为 θ 估计量. 通过 5000 次模拟运行计算 θ 的近似 0.95 置信区间的平均长度和经验覆盖概率, 其结果报告在表 2.5.2 中.

表 2.5.2　当维数 $p=1$ 时, 在不同的选择概率函数 $p(x,u)$ 和样本量 n 下 θ 的近似 0.95 置信区间的平均长度和经验覆盖概率

$p(x,u)$	n	平均长度			覆盖概率		
		WCEL	AEL	NA	WCEL	AEL	NA
$p_1(x,u)$	60	0.9409	0.9121	0.9282	0.9448	0.9354	0.9418
	100	0.7288	0.7153	0.7224	0.9456	0.9404	0.9436
	150	0.5947	0.5873	0.5912	0.9534	0.9492	0.9526
$p_2(x,u)$	60	0.9616	0.9291	0.9478	0.9398	0,9348	0.9366
	100	0.7457	0.7302	0.7386	0.9406	0.9362	0.9372
	150	0.6098	0.6012	0.6058	0.9530	0.9490	0.9514
$p_3(x,u)$	60	0.9780	0.9428	0.9631	0.9362	0.9286	0.9348
	100	0.7631	0.7461	0.7552	0.9380	0.9338	0.9362
	150	0.6254	0.6159	0.6208	0.9508	0.9476	0.9492

续表

$p(x,u)$	n	平均长度			覆盖概率		
		WCEL	AEL	NA	WCEL	AEL	NA
	60	0.9255	0.8974	0.9133	0.9442	0.9372	0.9402
$p_4(x,u)$	100	0.7163	0.7030	0.7101	0.9460	0.9412	0.9432
	150	0.5846	0.5773	0.5811	0.9544	0.9522	0.9552

由表 2.5.2 可得到下列三个结果. (i) WCEL 比 AEL 和 NA 有大的覆盖概率, 但有稍长的区间长度. (ii) 当 n 增大时覆盖概率增大而平均长度减小. (iii) 覆盖概率和平均长度依赖于选择概率函数. 这些结论与 Wang 等 (2004) 的结论基本一致.

2. 二维情形

考虑二维情形的部分线性模型 (2.1.1), 其中 $(\beta_1,\beta_2)=(0.5,1)$, (X_1,X_2) 服从二维正态分布 $N_2(0,0,1,1,0.6)$, ε 服从正态分布 $N(0,0.64)$. 选择概率取为

$$P(\delta=1|X,U)=\frac{\exp(0.85+0.07X_1-0.15X_2+0.02U)}{1+\exp(0.85+0.07X_1-0.15X_2+0.02U)},$$

其中 $\gamma=(0.85,0.07,-0.15,0.02)^{\mathrm{T}}$ 是参数. 使用广义线性模型方法估计 γ. 核函数取为 $K(x)=(15/16)(1-x^2)^2I(|x|\leqslant 1)$. 使用交叉验证方法选择最优带宽 h_{opt}, 且 M_n 取为 $\ln n+1$. 基于 IEL, CEL 和 NA 及样本量 $n=100,150,250$, 通过 400 次模拟运行计算 β_0 的置信域, 由 5000 次模拟运行计算 β_1 和 β_2 的置信区间的平均长度和相应的覆盖概率. 模拟结果展示在图 2.5.2 和表 2.5.3 中.

表 2.5.3 当维数 $p=2$ 时, 在不同的样本量 n 下 β_1 和 β_2 的近似 0.95 置信区间的平均长度和经验覆盖概率

β	n	平均长度				覆盖概率			
		IEL	CEL	NA($\hat\beta_{\mathrm{I}}$)	NA($\hat\beta_{\mathrm{C}}$)	IEL	CEL	NA($\hat\beta_{\mathrm{I}}$)	NA($\hat\beta_{\mathrm{C}}$)
	100	0.3747	0.3849	0.3552	0.3718	0.9076	0.9256	0.8968	0.9144
β_1	150	0.3073	0.3125	0.2954	0.3045	0.9266	0.9380	0.9168	0.9314
	250	0.2376	0.2403	0.2314	0.2362	0.9328	0.9384	0.9278	0.9332
	100	0.3758	0.3854	0.3563	0.3720	0.9146	0.9264	0.9024	0.9198
β_2	150	0.3070	0.3120	0.2951	0.3040	0.9278	0.9340	0.9174	0.9288
	250	0.2376	0.2404	0.2314	0.2362	0.9334	0.9376	0.9284	0.9344

图 2.5.2 表明 IEL 比 CEL 和 NA 给出小的置信域. 表 2.5.3 表明 IEL 比 CEL 有短的区间长度, 但有小的覆盖概率. IEL 和 CEL 比 NA($\hat\beta_{\mathrm{C}}$) 和 NA($\hat\beta_{\mathrm{I}}$) 有大的覆盖概率, 但有稍长的区间长度. 此外, 当 n 增大时经验覆盖概率都增大且区间长度减小.

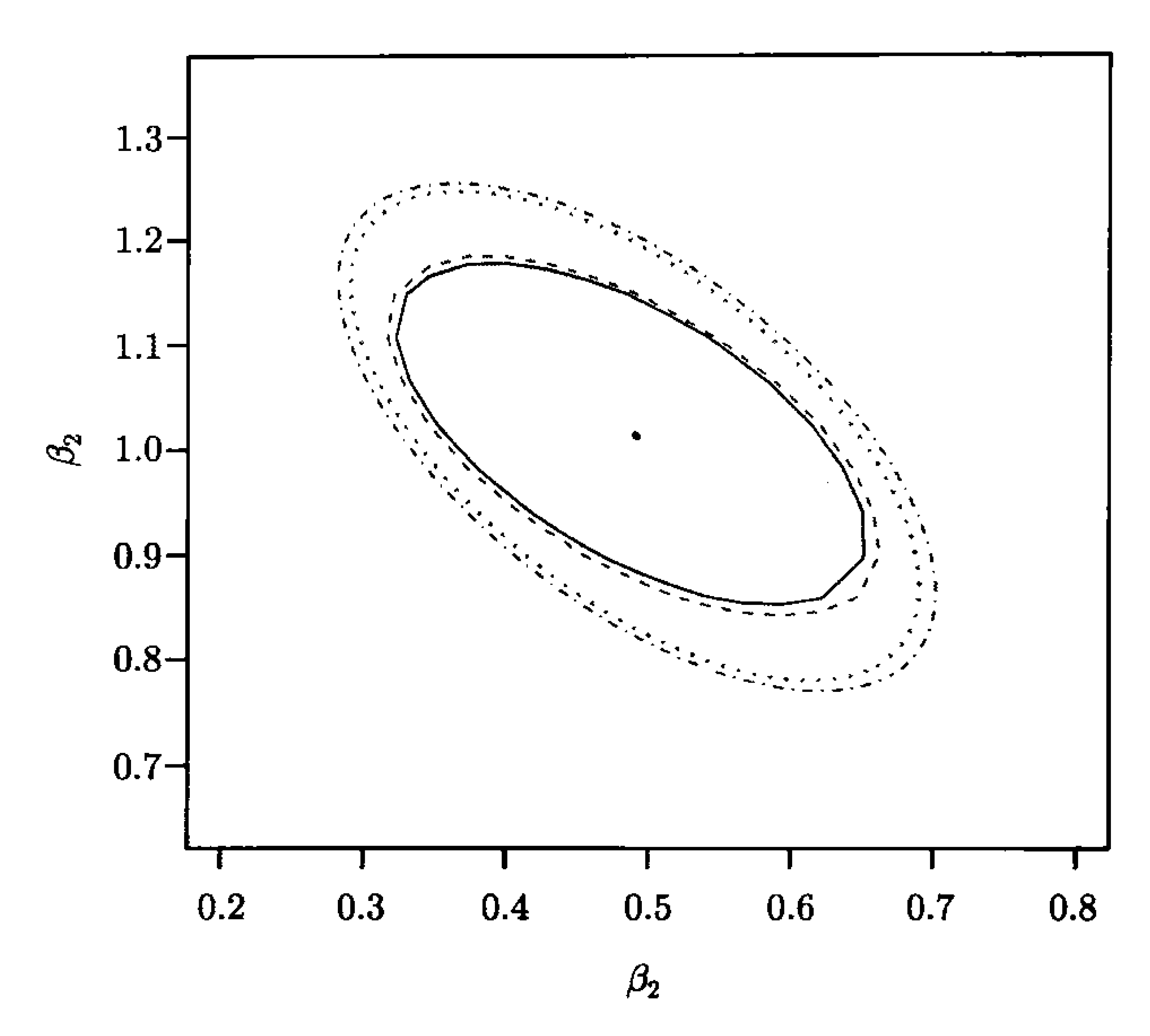


图 2.5.2　基于 IEL(实线)、CEL(虚线)、NA($\hat{\beta}_{\mathrm{I}}$)(点线) 和 NA($\hat{\beta}_{\mathrm{C}}$)(点虚线) 以及 $n = 100$ 和 $p = 2$, 参数 (β_1, β_2) 的近似 0.95 置信域

当 $n = 100$ 时, 基于 RAEL、EEL 和 NA, 我们用 400 次模拟运行计算了 $g(u)$ 的近似 0.95 置信带. 计算结果在图 2.5.3 中给出. 从图 2.5.3 可以看出, RAEL 比 EEL 和 NA 给出了狭窄的置信带.

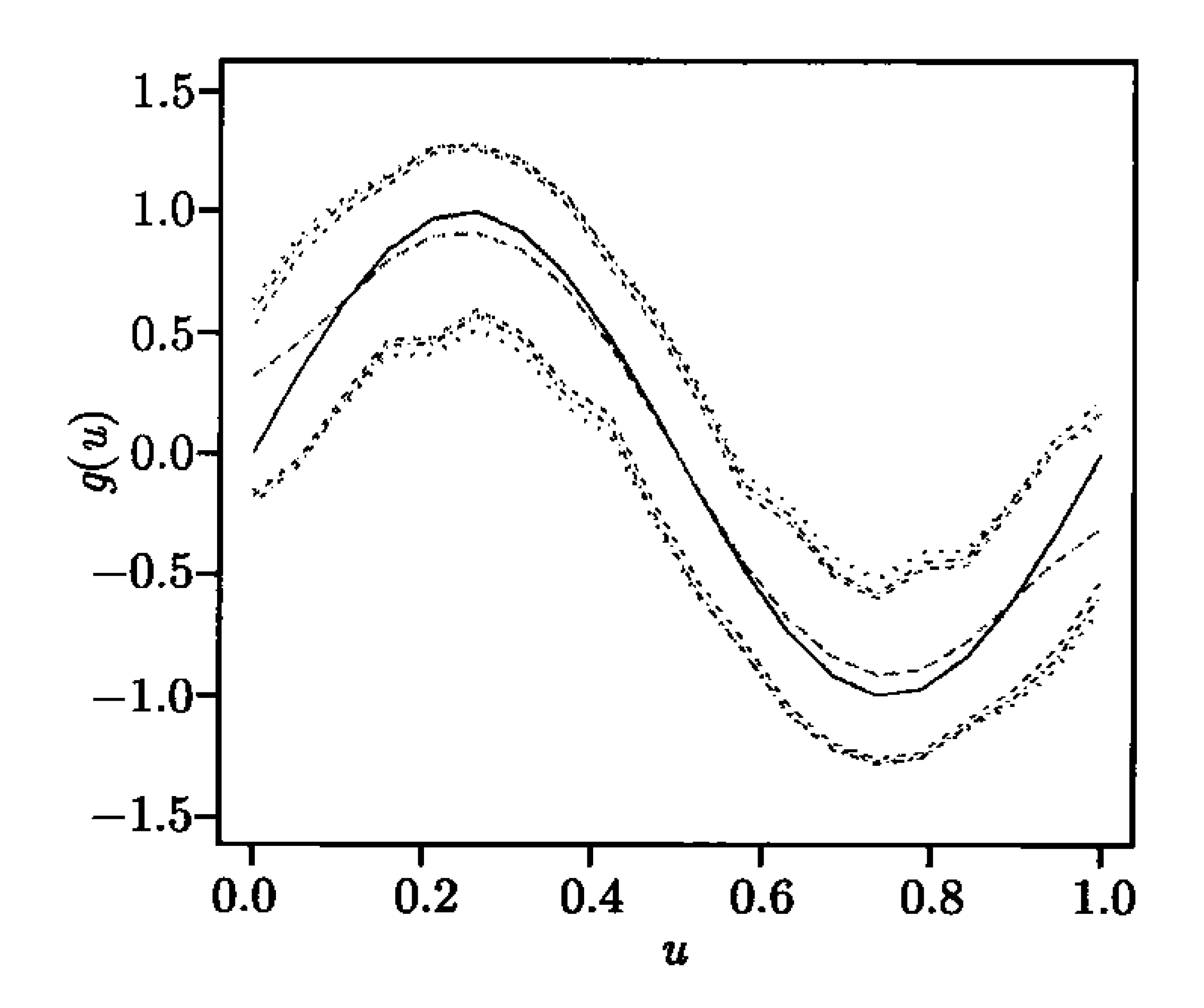


图 2.5.3　基于 RAEL(虚线)、EEL(点线) 和 NA(点虚线) 以及 $n = 100$ 和 $p = 2$, $g(u)$ 的近似 0.95 置信带. 实线表示 $g(u)$ 的真实曲线, 长虚线表示 $g(u)$ 的估计的曲线

我们用 5000 次运行计算了响应均值 θ 的近似 0.95 置信区间和相应的覆盖概率. 结果报告在表 2.5.4 中.

表 2.5.4 当维数 $p=2$ 时, 在不同的样本量 n 下 θ 的近似 0.95 置信区间的平均长度和经验覆盖概率

n	平均长度			覆盖概率		
	WCEL	AEL	NA	WCEL	AEL	NA
60	0.8665	0.8377	0.8545	0.9432	0.9350	0.9396
100	0.6708	0.6572	0.6648	0.9482	0.9416	0.9462
150	0.5479	0.5404	0.5445	0.9506	0.9478	0.9494

表 2.5.4 表明 WCEL 比 AEL 和 NA 有大的覆盖概率, 但有大的区间长度. 此外, 当 n 增大时覆盖概率都增大且区间长度减小.

2.5.6 定理的证明

为证明定理 2.5.1~ 定理 2.5.14, 首先给出下列引理.

引理 2.5.1 设条件 (C2.5.1)~(C2.5.5) 成立, 则对所有 $1\leqslant i\leqslant n$, 一致地有

$$E[\hat{\mu}_l(U_i)-\mu_l(U_i)]^2=O\left((nh)^{-1}\right)+O\left(h^4\right),\quad l=1,2.$$

证 仅证明 $l=1$ 的情况, 类似地可证明 $l=2$ 的情况. 利用 Cauchy-Schwarz 不等式可得

$$\begin{aligned}&E[\hat{\mu}_1(U_i)-\mu_1(U_i)]^2\\ \leqslant\ &2E\left[\sum_{j=1}^{n}\Omega_{nj}(U_i)\{Y_j-\mu_1(U_j)\}\right]^2\\ &+2E\left[\sum_{j=1}^{n}\Omega_{nj}(U_i)\{\mu_1(U_j)-\mu_1(U_i)\}\right]^2\\ \equiv\ &L_1+L_2.\end{aligned}\tag{2.5.30}$$

记 $r(u)=q(u)f(u)$,

$$\hat{r}(u)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{U_i-u}{h}\right)\delta_i.$$

使用非参数的规范证明方法可以证得

$$\sup_{0\leqslant t\leqslant 1}|\hat{r}(u)-r(u)|=O\left(n^{-1/5}\right),\quad \text{a.s..}$$

因此, 由条件 (C2.5.1) 可知, 存在正的常数 d_1 和 d_2 使得

$$\inf_{0\leqslant t\leqslant 1}r(u)\geqslant d_1,\qquad \inf_{0\leqslant t\leqslant 1}\hat{r}(u)\geqslant d_2,\quad \text{a.s..}$$

首先考虑 L_1. 注意到 $E[\delta_j(Y_j-\mu_1(U_j))]=0$, 因此

$$\begin{aligned} L_1 &\leqslant c(nh)^{-2}E\left[\sum_{j=1}^n K\left(\frac{U_j-U_i}{h}\right)\delta_j\{Y_j-\mu_1(U_j)\}\right]^2 \\ &\leqslant c(nh)^{-2}\sum_{j=1}^n E\left[K\left(\frac{U_j-U_i}{h}\right)\delta_j\{Y_j-\mu_1(U_j)\}\right]^2 \\ &\leqslant c(nh)^{-1}. \end{aligned} \tag{2.5.31}$$

下面处理 L_2. 记 $\psi(U_j,U_i)=K((U_j-U_i)/h)\delta_j[\mu_1(U_j)-\mu_1(U_i)]$. 使用条件 (C2.5.1), (C2.5.2) 和 (C2.5.4) 可得

$$|E[\psi(U_j,U_i)]|=O(h^3),\quad E[\psi^2(U_j,U_i)]=O(h^3).$$

因此可得

$$\begin{aligned} L_2 &\leqslant c(nh)^{-2}E\left[\sum_{j=1}^n\{\psi(U_j,U_i)-E[\psi(U_j,U_i)]\}\right]^2 \\ &\quad +c(nh)^{-2}\left\{\sum_{j=1}^n E[\psi(U_j,U_i)]\right\}^2 \\ &\leqslant cn^{-1}h+h^4. \end{aligned} \tag{2.5.32}$$

再结合 (2.5.30)~(2.5.32) 即证明了引理 2.5.1 的 $l=1$ 的情况.

引理 2.5.2　设条件 (C2.5.7)~(C2.5.9) 成立, 则对所有 $1\leqslant i\leqslant n$, 一致地有

$$E[\hat{p}(Z_i)-p(Z_i)]^2=O\left((nb^{p+1})^{-1}M_n^{p+1}\right)+O\left(b^{2r}\right)+o\left(n^{-1/2}\right),$$

其中 $Z_i=(X_i,U_i)$, $1\leqslant i\leqslant n$.

证　首先证明: 对所有 $1\leqslant i\leqslant n$, 一致地有

$$E\left[\frac{1}{C_n(Z_i)}\right]=O\left((nb^{p+1})^{-1}M_n^{p+1}\right)+o(n^{-1/2}), \tag{2.5.33}$$

其中

$$C_n(Z_i)=\max\left\{1,\ C_1\sum_{j\neq i}I[\|Z_j-Z_i\|\leqslant\rho b]\right\}.$$

注意到 $C_n(\cdot)$ 与 Spiegelman 和 Sacks(1980) 研究的 K_n^* 有关. 遵照 Spiegelman 和 Sacks(1980) 的证明方法或直接论证可得

$$E\left[\frac{1}{C_n(Z_i)}\bigg|Z_i=u\right]\leqslant\frac{c}{n\pi_n(u)},$$

其中

$$\pi_n(u) = P(||Z-u|| \leqslant \rho b).$$

由 (Spiegelman and Sacks, 1980) 中的引理 2 可知, 函数

$$z \longmapsto \int_{\{||z-u||\leqslant \rho b\}} \frac{\mu(\mathrm{d}u)}{\pi_n(u)}$$

关于 z 和 b 一致有界, 其中 $\mu(\cdot)$ 为 $Z=(X,U)$ 的分布. 注意到 $C_n(Z_i)\geqslant 1$, 可得

$$E\left[\frac{1}{C_n(Z_i)}\right] \leqslant 6\int_{||u||\leqslant M_n} \frac{\mu(\mathrm{d}u)}{n\pi_n(u)} + \int_{||u||>M_n} \mu(\mathrm{d}u) \equiv I + II. \tag{2.5.34}$$

为处理 I, 我们用有限多个以 u_i 为中心且 $\rho b/2$ 为半径的球 $S(u_i,\rho b/2)$ 覆盖 $\{||u||\leqslant M_n\}$. 所需要的这种球的数目显然为 $O(M_n^{p+1}(\rho b)^{-(p+1)})$. 如果可能, 在 $S(u_i,\rho b/2)$ 中选择 u. 容易看到 $S(u_i,\rho b/2)\subset S(u,\rho b)$. 因此,

$$\begin{aligned} I &\leqslant 6\sum_i \int_{S(u_i,\rho b/2)} \frac{\mu(\mathrm{d}u)}{n\pi_n(u)} \\ &\leqslant cM_n^{p+1}(\rho b)^{-(p+1)}n^{-1}\int_{S(u,\rho b)} \frac{\mu(\mathrm{d}u)}{\pi_n(u)} \\ &\leqslant c(nb^{p+1})^{-1}M_n^{p+1}. \end{aligned} \tag{2.5.35}$$

由条件 (C2.5.9) 可知 $II=o(n^{-1/2})$, 此式与 (2.5.34) 和 (2.5.35) 式结合即证明了 (2.5.33) 式.

下面证明引理 2.5.2. 记

$$\widetilde{W}_{nj}(z) = \widetilde{K}\left(\frac{Z_j-z}{b}\right) \Big/ \max\left\{1, \sum_{i=1}^n \widetilde{K}\left(\frac{Z_i-z}{b}\right)\right\}.$$

简单计算可得

$$\begin{aligned} E[\hat{p}(Z_i)-p(Z_i)]^2 \leqslant\ & 3E\left[\sum_{j=1}^n \widetilde{W}_{nj}(Z_i)\{\delta_j - p(Z_j)\}\right]^2 \\ &+3E\left[\sum_{j=1}^n \widetilde{W}_{nj}(Z_i)\{p(Z_j)-p(Z_i)\}\right]^2 \\ &+3E\left[\left(\sum_{j=1}^n \widetilde{W}_{nj}(Z_i) - 1\right)p(Z_i)\right]^2 \\ \equiv\ & I_1 + I_2 + I_3. \end{aligned} \tag{2.5.36}$$

由独立性和正交性, 并使用 (2.5.33) 式可得

$$\begin{aligned}I_1 &= 3\sum_{j=1}^{n} E\left\{\widetilde{W}_{nj}^2(Z_i)E[(\delta_j - p(Z_j))^2|Z_j]\right\}\\&\leqslant cE\left[\sum_{j=1}^{n}\widetilde{W}_{nj}^2(Z_i)\right] \leqslant cE\left[\frac{1}{C_n(Z_i)}\right]\\&= O\left(M_n^{p+1}(nb^{p+1})^{-1}\right) + o\left(n^{-1/2}\right).\end{aligned} \tag{2.5.37}$$

为处理 I_2, 定义 $\hat{p}(z)$ 的伪期望

$$\bar{p}_n(z) = \frac{\int p(u)\widetilde{K}((u-z)/b)\mu(\mathrm{d}u)}{\int \widetilde{K}((u-z)/b)\mu(\mathrm{d}u)}.$$

当 $\widetilde{K}$ 是自然核时, $\bar{p}_n(z)$ 与 Spiegelman 和 Sacks(1980) 的函数 $M_n(z)$ 一致. 因此可得

$$\begin{aligned}I_2 &\leqslant 6E\left[\sum_{j=1}^{n}\widetilde{W}_{nj}(Z_i)\{p(Z_j) - \bar{p}_n(Z_i)\}\right]^2 + 6E[\bar{p}_n(Z_i) - p(Z_i)]^2\\&\equiv I_{21} + I_{22}.\end{aligned}$$

由 $\bar{p}_n$ 的定义, 每个 $p(Z_j)$ 是关于 Z_i 的条件中心化, 这就导致不相关的和, 因此可采用 Cauchy-Schwarz 不等式. 由正交性和 (2.5.33) 式可得

$$\begin{aligned}I_{21} &\leqslant cE\left\{E\left[\left(\sum_{j=1}^{n}\widetilde{W}_{nj}(Z_i)\Big(p(Z_j) - \bar{p}_n(Z_i)\Big)\right)^2\middle|Z_i\right]\right\}\\&\leqslant cE\left\{\sum_{j=1}^{n}E\left[\widetilde{W}_{nj}^2(Z_i)\left(p(Z_j) - \bar{p}_n(Z_i)\right)^2\middle|Z_i\right]\right\}\\&\leqslant cE\left[\sum_{j=1}^{n}\widetilde{W}_{nj}^2(Z_i)\right] \leqslant cE\left[\frac{1}{C_n(Z_i)}\right]\\&= O\left((nb^{p+1})^{-1}M_n^{p+1}\right) + o\left(n^{-1/2}\right).\end{aligned}$$

由条件 (C2.5.7) 和 (C2.5.8), 容易证明 $I_{22} = O(b^{2r})$. 这就证明了

$$I_2 = O\left((nb^{p+1})^{-1}M_n^{p+1}\right) + o\left(n^{-1/2}\right) + O\left(b^{2r}\right). \tag{2.5.38}$$

现在处理 I_3. 记

$$B_n(z) = \max\left\{1, \sum_{j=1}^{n}\widetilde{K}\left(\frac{Z_j - z}{b}\right)\right\}.$$

由 (2.5.33) 式可得

$$\begin{aligned} I_3 &\leqslant cP\{B_n(Z_i)=1\} \\ &\leqslant cE\left[\frac{1}{B_n(Z_i)}\right] \leqslant cE\left[\frac{1}{C_n(Z_i)}\right] \\ &= O\left((nb^{p+1})^{-1}M_n^{p+1}\right) + o\left(n^{-1/2}\right). \end{aligned} \tag{2.5.39}$$

将 (2.5.37)~(2.5.39) 式代入 (2.5.36) 式即证得引理 2.5.2.

引理 2.5.3 设条件 (C2.5.1)~(C2.5.10) 成立, 则

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}_i(\beta_0) \xrightarrow{D} N(0,\Lambda),$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i(\beta_0)\hat{\eta}_i^{\mathrm{T}}(\beta_0) \xrightarrow{P} \Lambda,$$

$$\max_{1\leqslant i\leqslant n}||\hat{\eta}_i(\beta_0)|| = o_P(n^{1/2}),$$

其中 $\hat{\eta}_i(\beta_0)$ 取作 $\hat{\eta}_{i,\mathrm{C}}(\beta_0)$ 或 $\hat{\eta}_{i,\mathrm{I}}(\beta_0)$, $\Lambda = E\{\phi(X,U)[X-\mu_2(U)]^{\otimes 2}\varepsilon^2\}$. 此处当 $\hat{\eta}_i(\beta_0)$ 取作 $\hat{\eta}_{i,\mathrm{C}}(\beta_0)$ 时, $\phi(x,u)=p(x,u)$, 此时 $\Lambda = B$; 当 $\hat{\eta}_i(\beta_0)$ 取作 $\hat{\eta}_{i,\mathrm{I}}(\beta_0)$ 时, $\phi(x,u)=1/p(x,u)$, 此时 $\Lambda = D$.

证 (a) 首先考虑 $\hat{\eta}_{i,\mathrm{C}}(\beta_0)$. 我们仅证第一式, 类似地可以证明其他两式. 对 $1\leqslant i\leqslant n$, 记 $X_i^* = X_i-\mu_2(U_i)$, $\breve{\mu}_2(U_i)=\mu_2(U_i)-\hat{\mu}_2(U_i)$, $\breve{g}(U_i)=g(U_i)-\hat{g}(U_i)$. 简单计算可得

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}_{i,\mathrm{C}}(\beta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i X_i^*\varepsilon_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\breve{\mu}_2(U_i)\varepsilon_i \\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i X_i^*\breve{g}(U_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\breve{\mu}_2(U_i)\breve{g}(U_i) \\ &\equiv J_1+J_2+J_3+J_4. \end{aligned}$$

容易得到 $E(J_1)=0$, $\mathrm{cov}(J_1)=B$. 由中心极限定理可得 $J_1 \xrightarrow{D} N(0,B)$. 因此, 为证引理 2.5.3, 仅需证明 $J_l \xrightarrow{P} 0$, $l=2,3,4$. 由引理 2.5.2 可得

$$\begin{aligned} E(||J_2||^2) &\leqslant cn^{-1}\sum_{i=1}^{n}E(||\breve{\mu}_2(U_i)||^2) \\ &\leqslant c[(nh)^{-1}+h^4] \longrightarrow 0, \end{aligned}$$

$$\begin{aligned} E\big(||J_4||\big) &\leqslant \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{E[||\breve{\mu}_2(U_i)||^2]\right\}^{1/2}\left\{E[\breve{g}^2(U_i)]\right\}^{1/2} \\ &\leqslant c(nh^2)^{-1/2}+c\sqrt{n}h^4 \to 0. \end{aligned}$$

这就推出 $J_2 \xrightarrow{P} 0$, $J_4 \xrightarrow{P} 0$.

现在处理 J_3. 用 $J_{3,s}$ 和 $X_{is}^*(u)$ 分别表示 J_3 和 $X_i^*(u)$ 的第 s 个分量. 直接计算可得

$$\begin{aligned}E(J_{3,s}^2) \leqslant\ & 2n^{-1}E\left[\sum_{j=1}^n\left(\sum_{i=1}^n \Omega_{nj}(U_i)\delta_i X_{is}^*\right)\delta_j\varepsilon_j\right]^2 \\ & +2n^{-1}E\left[\sum_{i=1}^n \delta_i X_{is}^*\left(g(U_i)-\sum_{j=1}^n \delta_j\Omega_{nj}(U_i)g(U_j)\right)\right]^2 \\ \equiv\ & I+II. \end{aligned} \tag{2.5.40}$$

注意到 $E(\varepsilon_j|X_j,U_j)=0$, $E(\delta_i X_i^*|U_i)=0$. 仿照 (2.5.31) 式的证明可证得

$$I=O\left((nh)^{-1}\right)+o(1), \quad II=O\left(n^{-1}h+h^4\right)+o(1). \tag{2.5.41}$$

结合 (2.5.40) 和 (2.5.41) 式可得

$$E(J_{3,s}^2) \leqslant c(nh)^{-1}+ch^4+o(1) \to 0.$$

这就证明了 $J_3 \xrightarrow{P} 0$. 综合上述证明即证得对 $\hat{\eta}_{i,\mathrm{C}}(\beta_0)$ 的第一式.

(b) 现在考虑 $\hat{\eta}_{i,\mathrm{I}}(\beta_0)$. 直接计算可得

$$\hat{\eta}_{i,\mathrm{I}}(\beta_0)=\hat{\eta}_{i,\mathrm{W}}(\beta_0)+\left(1-\frac{\delta_i}{\hat{p}(Z_i)}\right)\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}(\hat{\beta}_{\mathrm{C}}-\beta_0), \tag{2.5.42}$$

其中

$$\hat{\eta}_{i,\mathrm{W}}(\beta_0)=\frac{\delta_i}{\hat{p}(Z_i)}\widetilde{X}_i(\widetilde{Y}_i-\widetilde{X}_i^{\mathrm{T}}\beta_0).$$

可以证明

$$\frac{1}{n}\sum_{i=1}^n\left(1-\frac{\delta_i}{\hat{p}(Z_i)}\right)\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}=o_P(1), \tag{2.5.43}$$

$$\hat{\beta}_{\mathrm{C}}-\beta_0=A^{-1}\cdot\frac{1}{n}\sum_{i=1}^n\hat{\eta}_{i,\mathrm{C}}(\beta_0). \tag{2.5.44}$$

由 (2.5.44) 和引理 2.5.3 的第一式可知 $\hat{\beta}_{\mathrm{C}}-\beta_0=O_P(n^{-1/2})$. 再结合 (2.5.42) 和 (2.5.43) 式可证得

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{\eta}_{i,\mathrm{I}}(\beta_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{\eta}_{i,\mathrm{W}}(\beta_0)+o_P(1),$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{i,\mathrm{I}}(\beta_0)\hat{\eta}_{i,\mathrm{I}}^{\mathrm{T}}(\beta_0)=\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{i,\mathrm{W}}(\beta_0)\hat{\eta}_{i,\mathrm{W}}^{\mathrm{T}}(\beta_0)+o_P(1),$$

$$\max_{1\leqslant i\leqslant n}\|\hat{\eta}_{i,\mathrm{I}}(\beta_0)\|=\max_{1\leqslant i\leqslant n}\|\hat{\eta}_{i,\mathrm{W}}(\beta_0)\|+o_P(n^{1/2}).$$

下面的证明类似于情况 $\hat{\eta}_{i,\mathrm{C}}(\beta_0)$ 的证明, 这里从略.

引理 2.5.4 设条件 (C2.5.1)~(C2.5.5) 成立, 方差 $\sigma_\varepsilon^2(u)$ 在点 u_0 处连续. 如果 $g(u_0)$ 是基准函数的真值, 则

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{E}}(g(u_0))-b(u_0)\xrightarrow{D}N(0,v^2(u_0)),$$

$$\frac{1}{nh}\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{E}}^2(g(u_0))\xrightarrow{P}v^2(u_0),$$

$$\max_{1\leqslant i\leqslant n}|\hat{\zeta}_{i,\mathrm{E}}(g(u_0))|=o_P((nh)^{1/2}).$$

此外, 如果用 $nh^2/\log n\to\infty$ 和 $nh^5\to 0$ 代替条件 (C2.5.5), 则偏差项 $b(u_0)$ 是 0, 且上述结果仍成立.

证 仅证明第一式, 其他两式可类似证明. 使用定义在 $\hat{L}_{\mathrm{R}}(g(u_0))$ 中的记号 $\hat{\zeta}_{i,\mathrm{E}}(g(u_0))$, 可得

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{E}}(g(u_0))-b(u_0)=S_1(u_0)+S_2(u_0)+S_3(u_0),\tag{2.5.45}$$

其中

$$S_1(u_0)=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}K\left(\frac{U_i-u_0}{h}\right)\delta_i\varepsilon_i,$$

$$S_2(u_0)=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}K\left(\frac{U_i-u_0}{h}\right)\delta_i[g(U_i)-g(u_0)]-b(u_0),$$

$$S_3(u_0)=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}K\left(\frac{U_i-u_0}{h}\right)\delta_iX_i^{\mathrm{T}}(\beta-\hat{\beta}_{\mathrm{C}}).$$

不难证明 $E[S_1(u_0)]=0$, $\mathrm{var}(S_1(u_0))=v^2(u_0)+o(1)$. 可以验证 $S_1(u_0)$ 满足 Cramér-Wold 定理的条件和 Lindeberg 条件 (Serfling, 1980), 因此可得

$$S_1(u_0)\xrightarrow{D}N(0,v^2(u_0)).\tag{2.5.46}$$

我们也可以证明 $\mathrm{var}(S_2(u_0))=o(1)$. 因此

$$S_2(u_0)\xrightarrow{P}0.\tag{2.5.47}$$

容易证明

$$\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{U_i-u_0}{h}\right)\delta_i X_i=O_P(1).$$

由 (2.5.44) 式可知 $\hat{\beta}_C-\beta=O_P(n^{-1/2})$. 因此可得 $S_3(u_0)=O_P(h^{1/2})$. 该式与 (2.5.45)~(2.5.47) 式结合即可证得引理 2.5.4.

引理 2.5.5　设条件 (C2.5.1)~(C2.5.5) 成立; 又设方差 $\sigma_\varepsilon^2(u)$ 在点 u_0 处连续, 核 $K(\cdot)$ 在 $[0,1]$ 是二次连续可微的, 则 $\hat{b}(u_0)\xrightarrow{P}b(u_0)$.

证　记 $\varphi_i(u_0)=[g(U_i)-g(u_0)]\delta_i K((U_i-u_0)/h)$, $\hat{\varphi}_i(u_0)=[\hat{g}(U_i)-\hat{g}(u_0)]\delta_i \cdot K((U_i-u_0)/h)$. 那么有

$$\begin{aligned}\hat{b}(u_0)-b(u_0)&=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}[\hat{\varphi}_i(u_0)-\varphi_i(u_0)]\\&\quad+\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}[\varphi_i(u_0)-(h/n)^{1/2}b(u_0)]\\&=H_1(u_0)+H_2(u_0).\end{aligned}$$

由条件 (C2.5.1)~(C2.5.5) 和 Taylor 公式可得

$$\begin{aligned}&\hat{\varphi}_i(u_0)-\varphi_i(u_0)\\&=\{[\hat{g}'(u_0)-g'(u_0)](U_i-u_0)+o_P(|U_i-u_0|^2)\}\delta_i K((U_i-u_0)/h).\end{aligned}$$

使用条件 (C2.5.4) 和 (C2.5.5) 可以证明

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(U_i-u_0)^l\delta_i K\left(\frac{U_i-u_0}{h}\right)=O_P(1),\quad l=1,2$$

及 $\hat{g}'(u_0)-g'(u_0)\xrightarrow{P}0$. 因此可得 $H_1(u_0)\xrightarrow{P}0$. 在 (2.5.47) 式中已证得 $H_2(u_0)\xrightarrow{P}0$. 这就完成了引理 2.5.5 的证明.

引理 2.5.6　设条件 (C2.5.1)~(C2.5.10) 成立. 如果 θ 是参数真值, 则

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{Y}_i-\theta)\xrightarrow{D}N(0,V),$$

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i-\theta)^2\xrightarrow{P}V,$$

$$\max_{1\leqslant i\leqslant n}|\hat{Y}_i|=o_P(n^{1/2}),$$

其中 V 在定理 2.5.10 中定义.

证 仅证第一式, 其他两式可类似证明. 仍使用引理 2.5.1 和引理 2.5.2 的记号, 并记 $Z_i=(X_i,U_i)$, $i=1,\cdots,n$. 直接计算可得

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{Y}_i-\theta)=S_1+S_2+S_3+S_4,$$

其中

$$S_1=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\delta_i\varepsilon_i}{p(Z_i)}+[\beta_0^{\mathrm{T}}X_i+g(U_i)-\theta]\right\},$$

$$S_2=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{\hat{p}(Z_i)}-\frac{1}{p(Z_i)}\right)\delta_i\varepsilon_i,$$

$$S_3=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{\delta_i}{\hat{p}(Z_i)}\right)X_i^{\mathrm{T}}(\hat{\beta}_{\mathrm{C}}-\beta_0),$$

$$S_4=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{\delta_i}{\hat{p}(Z_i)}\right)[\hat{g}(U_i)-g(U_i)].$$

因 $\sqrt{n}S_1$ 是 iid 随机变量之和, 故由中心极限定理可得

$$S_1\xrightarrow{D}N(0,V).$$

为证明第一式, 仅需证明 $S_l=o_P(1)$, $l=2,3,4$. 首先处理 S_2. 直接计算可得

$$\begin{aligned}S_2&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i\varepsilon_i[p(Z_i)-\hat{p}(Z_i)]}{p^2(Z_i)}\\&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i\varepsilon_i[p(Z_i)-\hat{p}(Z_i)]^2}{p^2(Z_i)\hat{p}(Z_i)}\\&\equiv S_{21}+S_{22}.\end{aligned}$$

简单计算有

$$\begin{aligned}S_{21}&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\delta_i\varepsilon_i}{p^2(Z_i)}\sum_{j=1}^{n}\widetilde{W}_{nj}(Z_i)[p(Z_i)-p(Z_j)]\right\}\\&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\delta_i\varepsilon_i}{p^2(Z_i)}\sum_{j=1}^{n}\widetilde{W}_{nj}(Z_i)[p(Z_j)-\delta_j]\right\}\\&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\delta_i\varepsilon_i}{p(Z_i)}\left[1-\sum_{j=1}^{n}\widetilde{W}_{nj}(Z_i)\right]\right\}\\&\equiv S_{211}+S_{212}+S_{213}.\end{aligned}$$

由独立性和正交性, 并使用 Cauchy-Schwarz 不等式可得

$$\begin{aligned}E(S_{211}^2) &\leqslant \frac{c}{n}\sum_{i=1}^{n}E\left\{E\left[\sum_{j=1}^{n}\widetilde{W}_{nj}(Z_i)(p(Z_i)-p(Z_j))^2\Big|Z_i\right]\right\}\\ &\leqslant \frac{c}{n}\sum_{i=1}^{n}E\left\{E\left[\sum_{j=1}^{n}\widetilde{W}_{nj}(Z_i)\|Z_i-Z_j\|^2\Big|Z_i\right]\right\}\\ &\leqslant ch^2 \to 0.\end{aligned}$$

这就证明了 $S_{211}=o_P(1)$.

为处理 S_{212}, 记 $e_i'=\delta_i\varepsilon_i/p^2(Z_i)$, $e_j''=p(Z_j)-\delta_j$. 由独立性和正交性, 并使用 (2.5.33) 式可得

$$\begin{aligned}E(S_{212}^2) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}E\left[\widetilde{W}_{nj}(Z_i)\widetilde{W}_{nl}(X_k)e_i'e_j''e_k'e_l''\right]\\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[\widetilde{W}_{nj}^2(Z_i)e_i'^2e_j''^2\right]\\ &\leqslant \frac{c}{n}\sum_{i=1}^{n}E\left[\sum_{j=1}^{n}\widetilde{W}_{nj}^2(Z_i)\right]\leqslant \frac{c}{n}\sum_{i=1}^{n}E\left[\frac{1}{C_n(Z_i)}\right]\\ &= O\left(M_n^{p+1}(nb^{p+1})^{-1}\right)+o\left(n^{-1/2}\right)\to 0,\end{aligned}$$

这就推出 $S_{212}=o_P(1)$.

对 S_{213}, 由 (2.5.39) 式可得

$$\begin{aligned}E(S_{213}^2) &= \frac{1}{n}\sum_{i=1}^{n}E\left\{E\left[\frac{\delta_i\varepsilon_i^2}{p^2(Z_i)}\Big|Z_i\right]\left(1-\sum_{j=1}^{n}\widetilde{W}_{nj}(Z_i)\right)^2\right\}\\ &\leqslant \frac{c}{n}\sum_{i=1}^{n}P\left\{B_n(Z_i)=1\right\}\\ &= O\left((nb^{p+1})^{-1}M_n^{p+1}\right)+o\left(n^{-1/2}\right)\to 0.\end{aligned}$$

从而得到 $S_{213}=o_P(1)$. 这就证明了 $S_{21}=o_P(1)$. 由引理 2.5.2 和大数定律, 容易证明 $S_{22}=o_P(1)$. 这就证明了 $S_2=o_P(1)$. 由 (2.5.43) 和 (2.5.44) 式可得 $S_3=o_P(1)$.

现在处理 S_4. 记

$$S_{41}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{\delta_i}{p(Z_i)}\right)\{\hat{g}(U_i)-g(U_i)\}.$$

由引理 2.5.1 和引理 2.5.2 可得 $S_4 = S_{41} + o_P(1)$. 为证明 $S_4 = o_P(1)$, 仅需证明 $S_{41} = o_P(1)$. 下面考虑 S_{41}. 仍记 $r(u) = q(u)f(u)$,

$$\eta_n(u) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{U_j - u}{h}\right)\delta_j\varepsilon_j,$$

$$\zeta_n(u) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{U_j - u}{h}\right)\delta_j\{g(U_j) - g(u)\}.$$

由 (2.5.44) 和引理 2.5.1 可得

$$\begin{aligned}
S_{41} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p(Z_i)}\right)\cdot\frac{\eta_n(U_i)}{r(U_i)} \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p(Z_i)}\right)\cdot\frac{\zeta_n(U_i)}{r(U_i)} \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p(Z_i)}\right)\hat{\mu}_2(U_i)^{\mathrm{T}}(\beta - \hat{\beta}_{\mathrm{C}}) \\
&\equiv S_{411} + S_{412} + S_{413} + o_P(1).
\end{aligned}$$

下面证明 $S_{41l} = o_P(1)$, $l = 1, 2, 3$. 首先处理 S_{411}. 直接计算可得

$$\begin{aligned}
E(S_{411}^2) &= \frac{1}{n^3h^2}E\left[\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{K((U_j - U_i)/h)}{r(U_i)}\left(1 - \frac{\delta_i}{p(Z_i)}\right)\delta_j\varepsilon_j\right]^2 \\
&\leqslant \frac{c}{n^3h^2}\sum_{j=1}^{n}\sum_{i=1}^{n}E\left[\frac{K((U_j - U_i)/h)}{r(U_i)}\left(1 - \frac{\delta_i}{p(Z_i)}\right)\right]^2 \\
&\leqslant c(nh)^{-1} \to 0.
\end{aligned}$$

这就证明了 $S_{411} = o_P(1)$.

现在处理 S_{412}. 直接计算可以得到

$$\begin{aligned}
S_{412}^2 &\leqslant \frac{2}{n}\left\{\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p(Z_i)}\right)\cdot\frac{\zeta_n(U_i) - E[\zeta_n(U_i)|U_i]}{r(U_i)}\right\}^2 \\
&\quad + \frac{2}{n}\left\{\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p(Z_i)}\right)\cdot\frac{E[\zeta_n(U_i)|U_i]}{r(U_i)}\right\}^2 \\
&\equiv S'_{412} + S''_{412}.
\end{aligned}$$

由独立性和正交性, 并使用 (2.5.33) 式、条件 (C2.5.1)~(C2.5.5) 和 (C2.5.7) 可得

$$\begin{aligned}E(S_{412}'^2) \leqslant & \frac{2}{n}\sum_{i=1}^{n} E\left[\left(1-\frac{\delta_i}{p(Z_i)}\right)^2 \cdot \frac{\{\zeta_n(U_i)-E[\zeta_n(U_i)|U_i]\}^2}{m^2(U_i)}\right] \\ & + \frac{2}{n}\sum_{i\neq j} E\left[\left(1-\frac{\delta_i}{p(Z_i)}\right)\left(1-\frac{\delta_j}{p(Z_j)}\right)\right. \\ & \left. \times \frac{\{\zeta_n(U_i)-E[\zeta_n(U_i)|U_i]\}\{\zeta_n(U_j)-E[\zeta_n(U_j)|U_j]\}}{r(U_i)r(U_j)}\right] \\ \leqslant & c(nh^{-1})^{-1} + c(nh^2)^{-1} + ch^4 \to 0.\end{aligned}$$

直接计算可得

$$\begin{aligned}E(S_{412}''^2) \leqslant & \frac{2}{n}\sum_{i=1}^{n} E\left\{\left(1-\frac{\delta_i}{p(Z_i)}\right) \cdot \frac{E[\zeta_n(U_i)|U_i]}{r(U_i)}\right\}^2 \\ \leqslant & ch^4 \to 0.\end{aligned}$$

因此 $S_{412} = o_P(1)$. 由 (2.5.33) 和 (2.5.34) 式可得 $S_{413} = o_P(1)$. 这就证明了 $S_{41} = o_P(1)$. 至此, 我们完成了引理 2.5.6 的第一式的证明.

下面转向定理的证明. 我们仅证明定理 2.5.1、定理 2.5.2 和定理 2.5.6~2.5.10. 定理 2.5.3~2.5.5 的证明类似于定理 2.5.1 和定理 2.5.2 的证明; 定理 2.5.11 和定理 2.5.12 的证明可参见文献 (Wang, 2009) 的定理 4 和定理 5 的证明; 定理 2.5.13 的证明可参见文献 (Liang et al., 2007) 的定理 2 的证明; 定理 2.5.14 的证明可参见文献 (Yang et al., 2009) 定理 1 的证明.

定理 2.5.1 的证　由 Lagrange 乘子法, 可将 $\hat{R}_{\mathrm{C}}(\beta_0)$ 表示为

$$\hat{R}_{\mathrm{C}}(\beta_0) = 2\sum_{i=1}^{n}\log\left(1+\lambda^{\mathrm{T}}\hat{\eta}_{i,\mathrm{C}}(\beta_0)\right), \tag{2.5.48}$$

其中 $\lambda = \lambda(\beta_0)$ 为 $d\times 1$ 向量, 它是下列方程的解:

$$\sum_{i=1}^{n}\frac{\hat{\eta}_{i,\mathrm{C}}(\beta_0)}{1+\lambda^{\mathrm{T}}\hat{\eta}_{i,\mathrm{C}}(\beta_0)} = 0. \tag{2.5.49}$$

由引理 2.5.3, 并使用文献 (Owen, 1990) 中 (2.14) 式的证明方法可以证得

$$||\lambda|| = O_P(n^{-1/2}). \tag{2.5.50}$$

对 (2.5.48) 式应用 Taylor 公式并使用引理 2.5.3 和 (2.5.50) 式可得

$$\hat{R}_{\mathrm{C}}(\beta_0) = 2\sum_{i=1}^{n}\left\{\lambda^{\mathrm{T}}\hat{\eta}_{i,\mathrm{C}}(\beta_0) - [\lambda^{\mathrm{T}}\hat{\eta}_{i,\mathrm{C}}(\beta_0)]^2/2\right\} + o_P(1). \tag{2.5.51}$$

由 (2.5.49) 式可以推出

$$0=\sum_{i=1}^{n}\frac{\hat{\eta}_{i,\mathrm{C}}(\beta_0)}{1+\lambda^{\mathrm{T}}\hat{\eta}_{i,\mathrm{C}}(\beta_0)}$$

$$=\sum_{i=1}^{n}\hat{\eta}_{i,\mathrm{C}}(\beta_0)-\sum_{i=1}^{n}\hat{\eta}_{i,\mathrm{C}}(\beta_0)\hat{\eta}_{i,\mathrm{C}}^{\mathrm{T}}(\beta_0)\lambda+\sum_{i=1}^{n}\frac{\hat{\eta}_{i,\mathrm{C}}(\beta_0)\left[\lambda^{\mathrm{T}}\hat{\eta}_{i,\mathrm{C}}(\beta_0)\right]^2}{1+\lambda^{\mathrm{T}}\hat{\eta}_{i,\mathrm{C}}(\beta_0)}.$$

结合引理 2.5.3 和 (2.5.50) 式证明了

$$\sum_{i=1}^{n}[\lambda^{\mathrm{T}}\hat{\eta}_{i,\mathrm{C}}(\beta_0)]^2=\sum_{i=1}^{n}\lambda^{\mathrm{T}}\hat{\eta}_{i,\mathrm{C}}(\beta_0)+o_P(1),$$

$$\lambda=\left(\sum_{i=1}^{n}\hat{\eta}_{i,\mathrm{C}}(\beta_0)\hat{\eta}_{i,\mathrm{C}}^{\mathrm{T}}(\beta_0)\right)^{-1}\sum_{i=1}^{n}\hat{\eta}_{i,\mathrm{C}}(\beta_0)+o_P(n^{-1/2}).$$

因此, 利用 (2.5.51) 可得

$$\hat{R}_{\mathrm{C}}(\beta_0)=\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}_{i,\mathrm{C}}(\beta_0)\right)^{\mathrm{T}}B^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}_{i,\mathrm{C}}(\beta_0)\right)+o_P(1).$$

再利用引理 2.5.3 即可完成定理 2.5.1 的证明.

定理 2.5.2 的证 由 (2.5.44) 和引理 2.5.3 立刻推出.

定理 2.5.6 的证 由引理 2.5.4, 并类似于定理 2.5.1 的证明可得到

$$\hat{\mathcal{R}}_{\mathrm{E}}(g(u_0))=\left(\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{E}}(g(u_0))\right)^2\Bigg/\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{E}}^2(g(u_0))+o_P(1). \tag{2.5.52}$$

因此, 由 (2.5.52) 式和引理 2.5.4 即证得定理 2.5.6.

定理 2.5.7 的证 直接计算可得

$$\sqrt{nh}[\hat{g}(u_0)-g(u_0)]=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{E}}(g(u_0))/r(u_0)+o_P(1). \tag{2.5.53}$$

利用 (2.5.53) 式和引理 2.5.4 即可完成定理 2.5.7 的证明.

定理 2.5.8 的证 注意到

$$\begin{aligned}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{R}}(g(u_0))=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{E}}(g(u_0))-b(u_0)\\&-\left[\hat{b}(u_0)-b(u_0)\right],\end{aligned} \tag{2.5.54}$$

$$\begin{aligned}\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{R}}^2(g(u_0))=&\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{E}}^2(g(u_0))+\sum_{i=1}^{n}\hat{\varphi}_i^2(u_0)\\&-2\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{E}}(g(u_0))\hat{\varphi}_i(u_0),\end{aligned} \tag{2.5.55}$$

其中 $\hat{\varphi}_i(u_0)$ 在引理 2.5.5 中定义. 此外, 容易证明

$$\frac{1}{nh}\sum_{i=1}^{n}\hat{\varphi}_i^2(u_0) \xrightarrow{P} 0,$$

$$\frac{1}{nh}\sum_{i=1}^{n}\hat{\zeta}_{i,\mathrm{E}}(g(u_0))\hat{\varphi}_i(u_0) \xrightarrow{P} 0.$$

利用 (2.5.54) 和 (2.5.55) 式以及引理 2.5.4 即可完成定理 2.5.8 的证明.

定理 2.5.9 的证 由引理 2.5.6, 并类似于定理 2.5.1 的证明可以证明定理 2.5.9, 此处省略.

定理 2.5.10 的证 由 (2.5.23) 式可得

$$\sqrt{n}(\hat{\theta}-\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{Y}_i-\theta).$$

利用引理 2.5.6 即可完成定理 2.5.10 的证明.

2.6 纵向数据分析

2.6.1 模型

纵向数据部分线性模型依误差项的性质、协变量和响应变量随时间或个体变化的情形以及是否包含滞后因变量等假定的不同而形式各异. 一类重要的纵向数据部分线性模型具有形式

$$Y_{ij} = \beta_0^{\mathrm{T}}X_{ij} + g(U_{ij}) + \varepsilon_{ij}, \quad i=1,\cdots,n,\ j=1,\cdots,m_i, \tag{2.6.1}$$

其中 $(X_{ij}^{\mathrm{T}}, U_{ij})^{\mathrm{T}} \in \mathbf{R}^p\times\mathbf{R}$ 为协变量, β_0 是相应于个体水平协变量的未知 p 维参数向量, $g(\cdot)$ 是一元未知光滑函数. ε_{ij} 是均值为 0 的随机误差. 记 $\varepsilon_i = (\varepsilon_{i1},\varepsilon_{i2},\cdots,\varepsilon_{im_i})^{\mathrm{T}}$. 假定 $\varepsilon_i,\cdots,\varepsilon_n$ 相互独立, 而当 $j_1 \neq j_2$ 时, ε_{ij_1} 和 ε_{ij_2} 可能相关, $j_1, j_2 = 1,2,\cdots,m_i$. 按照 Lin 和 Carroll(2001a), 假定 $E(Y_{ij}|X_i,U_i) = E(Y_{ij}|X_{ij},U_{ij})$, 其中 $X_i = (X_{i1},\cdots,X_{im_i})^{\mathrm{T}}$ 和 $U_i = (U_{i1},\cdots,U_{im_i})^{\mathrm{T}}$ 表示来自第 i 个个体观察的协变量. 同理, $E(Y_{ij}|U_i) = E(Y_{ij}|U_{ij})$ 和 $E(X_{ij}|U_i) = E(X_{ij}|U_{ij})$, 并分别用 $g_1(U_{ij})$ 和 $g_2(U_{ij})$ 表示.

本节用 $V_i = V_i(X_i,U_i)$ 表示作业协方差阵, $\Sigma_i = \mathrm{var}(Y_i|X_i,U_i)$ 为 Y_i 的真协方差阵, 其中 $Y_i = (Y_{i1},\cdots,Y_{im_i})^{\mathrm{T}}$. 全节假定 V_i 可依赖于不同于 β_0 的有限维讨厌参数 τ. 对某个 τ^*, 当 $\Sigma_i = V_i(\tau^*)$ 时, 可以达到半参数有效估计.

由于对纵向数据部分线性模型进行研究, 不仅要考虑部分线性模型的半参数复杂性, 而且涉及纵向数据组内相关性. 因此, 与通常独立数据下部分线性模型的研

究相比较而言, 研究纵向数据下部分线性模型的大样本性质更加困难. 本节主要研究模型 (2.6.1) 中兴趣参数的估计和置信域构造问题.

2.6.2 后移算法

Zeger 和 Diggle(1994) 提出的后移 (backfitting) 算法可以用来构造参数向量和未知函数的估计量. 具体步骤如下. 给定第 i 个个体观测向量 Y_i 的协方差阵 V_i, 迭代过程由下面三个步骤组成:

步骤 1. 给定第 k 次迭代的估计值 $\hat{\beta}^{[k]}$, $g(U_{ij})$ 的估计值 $\hat{g}^{[k]}(U_{ij})$ 由下式计算:

$$\hat{g}(U_{ij};\hat{\beta}^{[k]}) = \hat{g}_1(U_{ij}) - \hat{g}_2^{\mathrm{T}}(U_{ij})\hat{\beta}^{[k]},$$

其中 $\hat{g}_1(U_{ij})$ 和 $\hat{g}_2(U_{ij})$ 分别是 $g_1(U_{ij})$ 和 $g_2(U_{ij})$ 的非参数核估计.

步骤 2. 给定 $\hat{g}^{[k]}(U_{ij})$, 记 $s_{ij}^{[k]} = Y_{ij} - \hat{g}^{[k]}(U_{ij})$. 由线性模型 $s_{ij}^{[k]} = \beta_0^{\mathrm{T}} X_{ij} + \varepsilon_{ij}$ 的广义最小二乘法可得 β_0 的估计

$$\hat{\beta}^{[k+1]} = \left(\sum_{i=1}^{n} X_i^{\mathrm{T}} V_i^{-1} X_i\right)^{-1} \sum_{i=1}^{n} X_i^{\mathrm{T}} V_i^{-1} s_i^{[k]},$$

其中 $s_i^{[k]} = (s_{i1}^{[k]}, \cdots, s_{im_i}^{[k]})^{\mathrm{T}}$.

步骤 3. 重复步骤 1 和 2 直到收敛. 最后得到 β_0 的估计量, 记为 $\hat{\beta}_{\mathrm{B}}$. $g(u)$ 的估计量定义为

$$\hat{g}_{\mathrm{B}}(u) = \hat{g}_1(u) - \hat{\beta}_{\mathrm{B}}^{\mathrm{T}} \hat{g}_2(u). \tag{2.6.2}$$

2.6.3 profile 核估计方程

Lin 和 Carroll(2001a) 提出了 profile 核估计方程方法来估计 β_0 和 $g(\cdot)$. 对于给定的 β, 使用非参数回归函数的核估计方法得到 $g(u)$ 的估计

$$\hat{g}(u;\beta) = \hat{g}_1(u) - \beta^{\mathrm{T}} \hat{g}_2(u).$$

其中 $\hat{g}_1(u)$ 和 $\hat{g}_1(u)$ 分别为 $g_1(u)$ 和 $g_2(u)$ 的核估计. 通过求解 profile 核广义估计方程

$$\sum_{i=1}^{n} \frac{\partial [X_i\beta + \hat{g}(U_i;\beta)]^{\mathrm{T}}}{\partial \beta} V_i^{-1} \{Y_i - [X_i\beta + \hat{g}(U_i;\beta)]\} = 0,$$

可得到 β_0 的 profile 核估计

$$\hat{\beta}_{\mathrm{P}} = \left(\sum_{i=1}^{n} \widetilde{X}_i^{\mathrm{T}} V_i^{-1} \widetilde{X}_i\right)^{-1} \sum_{i=1}^{n} \widetilde{X}_i^{\mathrm{T}} V_i^{-1} \widetilde{Y}_i, \tag{2.6.3}$$

其中 V_i 为作业协方差矩阵, $\widetilde{Y}_i = Y_i - \hat{g}_1(U_i)$, $\widetilde{X}_i = X_i - \hat{g}_2(U_i)$. 用 $\hat{\beta}_{\mathrm{P}}$ 代替 (2.6.2) 中的 $\hat{\beta}_{\mathrm{B}}$ 即可得到 $g(u)$ 的核估计, 记为 $\hat{g}_{\mathrm{P}}(u)$.

Lin 和 Carroll(2001a) 证明了以下结论: 只有欠光滑回归函数 $g(\cdot)$ 或者完全忽视组内相关结构 $(V_i = I)$ 的情况下, 这种传统的 profile 核方法才能导致 β_0 的 $\sqrt{n}$ 相合估计. 他们也推导了 β_0 的半参数有效性分值函数. 研究表明, 与独立数据不同, profile 核方法产生的 β_0 的估计不是半参数有效的.

需要指出, Fan 和 Li(2004) 首先使用局部多项式拟合方法估计非参数函数 $g(\cdot)$, 然后在作业独立协方差矩阵假定下, 使用 profile 最小二乘估计法构造参数分量 β_0 的估计, 并证明了估计量的渐近正态性, 同时也给出了渐近方差的相合估计.

2.6.4　样条逼近

本节的估计方法是基于基逼近 (Huang et al., 2007). 基本思想是用基底扩张逼近函数 $g(\cdot)$, 然后利用常规的广义估计方程 (GEE). 设 $g(u)$ 可以用样条函数逼近以致

$$g(u) \approx \sum_{k=1}^{K_n} \gamma_{k0} B_k(u) = \gamma_0^{\mathrm{T}} B(u), \tag{2.6.4}$$

其中 $\{B_k(u)\}_{k=1}^{K_n}$ 是一个 B 样条基, $B=(B_1(u), \cdots, B_{K_n}(u))^{\mathrm{T}}, \gamma_0=(\gamma_{10}, \cdots, \gamma_{K_n0})^{\mathrm{T}}$. 当 $g(u)$ 为光滑函数时, 这样的样条逼近存在. 事实上, 当 $g(u)$ 连续时, 可以选择样条函数满足 $\sup_u |g(u) - \gamma_0^{\mathrm{T}} B(u)| \to 0$, $K_n \to \infty$(de Boor, 2001). 记 $Z_{ij} = B(U_{ij})$, $Z_i = (Z_{i1}, \cdots, X_{im_i})^{\mathrm{T}}$. 设 $\hat{\beta}_{\mathrm{SA}}$ 和 $\hat{\gamma}_{\mathrm{SA}}$ 为下列极小化问题的解:

$$\sum_{i=1}^{n} [Y_i - (X_i\beta + Z_i\gamma)]^{\mathrm{T}} V_i^{-1} [Y_i - (X_i\beta + Z_i\gamma)] = \min_{\beta,\gamma}!$$

或等价地, 它们是下列估计方程组的解:

$$\begin{cases} \displaystyle\sum_{i=1}^{n} X_i V_i^{-1} [Y_i - (X_i\beta + Z_i\gamma)] = 0, \\ \displaystyle\sum_{i=1}^{n} Z_i V_i^{-1} [Y_i - (X_i\beta + Z_i\gamma)] = 0. \end{cases} \tag{2.6.5}$$

因此, 用 $\hat{\beta}_{\mathrm{SA}}$ 估计模型的参数分量, 用 $\hat{g}_{\mathrm{SA}}(u) = \hat{\gamma}_{\mathrm{SA}}^{\mathrm{T}} B(u)$ 估计模型的非参数分量. 记 $\xi_i = (X_i, Z_i)$, 由 (2.6.5) 式容易得到

$$\begin{pmatrix} \hat{\beta}_{\mathrm{SA}} \\ \hat{\gamma}_{\mathrm{SA}} \end{pmatrix} = \left(\sum_{i=1}^{n} \xi_i^{\mathrm{T}} V_i^{-1} \xi_i \right)^{-1} \sum_{i=1}^{n} \xi_i^{\mathrm{T}} V_i^{-1} Y_i. \tag{2.6.6}$$

用 $\mathcal{X}$ 和 $\mathcal{Z}$ 分别表示 X_{ij} 和 Z_{ij} 的全体, 并记 $R(u) = g(u) - \gamma_0^{\mathrm{T}} B(u)$. 经过计算可以得到

$$E\left[\left. \begin{pmatrix} \hat{\beta}_{\mathrm{SA}} \\ \hat{\gamma}_{\mathrm{SA}} \end{pmatrix} \right| \mathcal{X}, \mathcal{Z} \right] = \begin{pmatrix} \beta_0 \\ \gamma_0 \end{pmatrix} + \left(\sum_{i=1}^{n} \xi_i^{\mathrm{T}} V_i^{-1} \xi_i \right)^{-1} \sum_{i=1}^{n} \xi_i^{\mathrm{T}} V_i^{-1} R(U_i),$$

$$\operatorname{var}\left(\left(\begin{matrix}\hat{\beta}_{\mathrm{SA}}\\ \hat{\gamma}_{\mathrm{SA}}\end{matrix}\right)\middle|\mathcal{X},\mathcal{Z}\right)=\left(\sum_{i=1}^{n}\xi_i^{\mathrm{T}}V_i^{-1}\xi_i\right)^{-1}\sum_{i=1}^{n}\xi_i^{\mathrm{T}}V_i^{-1}\left(\sum_{i=1}^{n}\xi_i^{\mathrm{T}}V_i^{-1}\xi_i\right)^{-1}.$$

从上述两个公式容易推出 $\hat{g}_{\mathrm{SA}}(u)$ 的条件均值和方差.

下面给出一些表达式, 以便进一步进行分析. 记

$$\sum_{i=1}^{n}\xi_i^{\mathrm{T}}V_i^{-1}\xi_i=\left(\begin{matrix}\sum_{i=1}^{n}X_i^{\mathrm{T}}V_i^{-1}X_i & \sum_{i=1}^{n}X_i^{\mathrm{T}}V_i^{-1}Z_i\\ \sum_{i=1}^{n}Z_i^{\mathrm{T}}V_i^{-1}X_i & \sum_{i=1}^{n}Z_i^{\mathrm{T}}V_i^{-1}Z_i\end{matrix}\right)\equiv\left(\begin{matrix}H_{11} & H_{12}\\ H_{21} & H_{22}\end{matrix}\right).$$

由熟知的分块矩阵的逆矩阵形式可以推出

$$\left(\begin{matrix}H_{11} & H_{12}\\ H_{21} & H_{22}\end{matrix}\right)^{-1}=\left(\begin{matrix}H_{11\cdot 2}^{-1} & -H_{11\cdot 2}^{-1}H_{12}H_{22}^{-1}\\ -H_{22\cdot 1}^{-1}H_{21}H_{11}^{-1} & H_{22\cdot 1}^{-1}\end{matrix}\right),$$

其中 $H_{11\cdot 2}=H_{11}-H_{12}H_{22}^{-1}H_{21}$, $H_{22\cdot 1}=H_{22}-H_{21}H_{11}^{-1}H_{12}$. 因此, 估计量 $\hat{\beta}_{\mathrm{SA}}$ 可表示为

$$\hat{\beta}_{\mathrm{SA}}=H_{11\cdot 2}^{-1}\sum_{i=1}^{n}(X_i-Z_iH_{22}^{-1}H_{21})^{\mathrm{T}}V_i^{-1}Y_i.$$

于是, $\hat{\beta}_{\mathrm{SA}}$ 的条件均值和条件方差为

$$E(\hat{\beta}_{\mathrm{SA}}|\mathcal{X},\mathcal{U})=\beta_0+H_{11\cdot 2}^{-1}\sum_{i=1}^{n}(X_i-Z_iH_{22}^{-1}H_{21})^{\mathrm{T}}V_i^{-1}R(U_i),$$

$$\operatorname{var}(\hat{\beta}_{\mathrm{SA}}|\mathcal{X},\mathcal{U})=H_{11\cdot 2}^{-1}\sum_{i=1}^{n}[(X_i-Z_iH_{22}^{-1}H_{21})^{\mathrm{T}}V_i^{-1}\Sigma_iV_i^{-1}(X_i-Z_iH_{22}^{-1}H_{21})]H_{11\cdot 2}^{-1}.$$

特别地, 当作业协方差阵等于真协方差阵 ($V_i=\Sigma_i$) 时, $\operatorname{var}(\hat{\beta}_{\mathrm{SA}}|\mathcal{X},\mathcal{U})=H_{11\cdot 2}^{-1}$. 使用同样的方法可得

$$\operatorname{var}(\hat{\gamma}_{\mathrm{SA}}|\mathcal{X},\mathcal{U})=H_{22\cdot 1}^{-1}\sum_{i=1}^{n}[(Z_i-X_iH_{11}^{-1}H_{12})^{\mathrm{T}}V_i^{-1}\Sigma_iV_i^{-1}(Z_i-X_iH_{11}^{-1}H_{12})]H_{22\cdot 1}^{-1}.$$

注意到 $\operatorname{var}(\hat{g}_{\mathrm{SA}}(u)|\mathcal{X},\mathcal{U})=B^{\mathrm{T}}(u)\operatorname{var}(\hat{\gamma}_{\mathrm{SA}}|\mathcal{X},\mathcal{U})B(u)$, 于是, 利用此式可以得到 var $(\hat{g}_{\mathrm{SA}}(u)|\mathcal{X},\mathcal{U})$ 的表达式.

Huang 等 (2007) 证明了: 当 $V_i=\Sigma_i$ 时, $\operatorname{var}\left(\hat{\beta}_{\mathrm{SA}}|\mathcal{X},\mathcal{Z}\right)$ 和 $\operatorname{var}(\hat{g}_{\mathrm{SA}}(u)|\mathcal{X},\mathcal{Z})$ 达到最小, 其中 u 为 U_{ij} 的支撑集中的点. 这个结果说明: 在这类基于样条的扩展 GEE 估计量中, 最有效的估计量是人们正确指定协方差结构. 这是一个有限样本下的结果, 即对任何固定的 n 和 m_i 都成立, 不牵涉渐近分析. Huang 等 (2007) 也

证明了: 即使错误指定协方差阵, 所提出的估计量仍具有相合性和渐近正态性. 通过证明他们的估计达到半参数信息界, Huang 等 (2007) 在更强的意义下确实建立了估计参数分量的有效性. 仅通过假定条件矩限制代替严格的多维 Gauss 误差假定, 即可得到提出的估计量的半参数有效性.

2.6.5 经验似然

下文假定 m_i 是有界的. 该假定意味着总的样本容量 N 与个体数 n 是同阶的量. 进一步假设 $\{U_{ij}, 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m_i\}$ 是 iid 的, 其共同的密度 $f(\cdot)$ 是 Lebesgue 可测的. 虽然时间设计点列的 iid 假定理论上仅适合随机设计情况, 但本节的结果和证明可以作适当修改后同样适合固定设计情况. 本节内容主要取材于文献 (Xue and Zhu, 2005).

1. *回归系数的经验似然*

在 (2.6.1) 式两边求给定 U_{ij} 下的条件期望, 并与 (2.6.1) 式两边分别相减可得

$$Y_{ij} - E(Y_{ij}|U_{ij}) = \beta_0^{\mathrm{T}}[X_{ij} - E(X_{ij}|U_{ij})] + \varepsilon_{ij}.$$

为构造 β_0 的经验似然比函数, 引入辅助随机变量

$$Z_i(\beta_0) = \sum_{j=1}^{m_i} w(U_{ij})\check{X}_{ij}(\check{Y}_{ij} - \beta_0^{\mathrm{T}}\check{X}_{ij}), \quad i = 1, \cdots, n,$$

其中 $\check{Y}_{ij} = Y_{ij} - g_1(U_{ij})$, $\check{X}_{ij} = X_{ij} - g_2(U_{ij})$, $w(\cdot)$ 是一个有界的非负权函数, 且具有紧支撑 $[a, b] \subset [0, 1]$. 这里 $g_1(u) = E[Y_{ij}|U_{ij} = u]$, $g_2(u) = E[X_{ij}|U_{ij} = u]$. 注意到 $E[Z_i(\beta)] = 0$ 当且仅当 $\beta = \beta_0$. 使用这个信息, 可以定义经验对数似然比函数 $l(\beta_0)$. 可证明 $l(\beta_0)$ 渐近于自由度为 p 的卡方分布. 然而, $l(\beta_0)$ 不能直接应用于 β_0 的统计推断, 因为它包含两个未知函数 $g_1(\cdot)$ 和 $g_2(\cdot)$. 解决这个问题的一个自然想法是在 $l(\beta_0)$ 中用两个估计量分别代替它们. 对于固定的点 $u \in [0, 1]$, 利用核估计法定义 $g_1(u)$ 和 $g_2(u)$ 的估计量分别为

$$\hat{g}_1(u) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} W_{ij}(u)Y_{ij},$$

$$\hat{g}_2(u) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} W_{ij}(u)X_{ij},$$

其中

$$W_{ij}(u) = K\left(\frac{U_{ij} - u}{h}\right) \Big/ \sum_{k=1}^{n}\sum_{l=1}^{m_k} K\left(\frac{U_{kl} - u}{h}\right), \tag{2.6.7}$$

$h=h_n$ 是带宽, $K(\cdot)$ 是核函数. 因此, 用 $\hat{g}_1(U_{ij})$ 和 $\hat{g}_2(U_{ij})$ 分别代替 $Z_i(\beta_0)$ 中的 $g_1(U_{ij})$ 和 $g_2(U_{ij})$ 可以得到 $Z_i(\beta_0)$ 的估计量 $\hat{Z}_i(\beta_0)$, 即

$$\hat{Z}_i(\beta_0)=\sum_{j=1}^{m_i}w(U_{ij})\widetilde{X}_{ij}(\widetilde{Y}_{ij}-\beta_0^{\mathrm{T}}\widetilde{X}_{ij}),\quad i=1,\cdots,n,$$

其中 $\widetilde{X}_{ij}=X_{ij}-\hat{g}_2(U_{ij})$, $\widetilde{Y}_{ij}=Y_{ij}-\hat{g}_1(U_{ij})$. 那么, β_0 的经验对数似然比函数定义为

$$\hat{l}(\beta_0)=-2\max\left\{\sum_{i=1}^{n}\log(np_i)\Big|p_i\geqslant 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\hat{Z}_i(\beta_0)=0\right\}.\qquad(2.6.8)$$

为了给出本节的主要结果, 首先列出所需要的条件.

(C2.6.1) 带宽满足 $h=h_0N^{-1/5}$, 其中 $h_0>0$ 为某个常数.

(C2.6.2) 核 $K(\cdot)$ 是对称的概率密度函数, 且在它的支撑集 $[-1,1]$ 上有界变差.

(C2.6.3) $\sup\limits_{x,0\leqslant u\leqslant 1}E(\varepsilon_{ij}^4|X_{ij}=x,U_{ij}=u)<\infty$, $\sup\limits_{0\leqslant u\leqslant 1}E(\|X_{ij}\|^2|U_{ij}=u)<\infty$, $E(\|X_{ij}\|^4)<\infty$, $i=1,\cdots,n$, $j=1,\cdots,m_i$.

(C2.6.4) 密度函数 $f(u)$ 在 $(0,1)$ 上是连续可微的, 且存在正的常数 c_1 和 c_2, 使得对一切 $u\in[0,1]$, 有 $c_1\leqslant f(u)\leqslant c_2$.

(C2.6.5) $g_1(u)$ 和 $g_2(u)$ 的每个分量在 $(0,1)$ 上是二次连续可微的.

(C2.6.6) A 是正定矩阵, 这里 A 在定理 2.6.2 中定义.

(C2.6.7) 方差函数 $\sigma_\varepsilon^2(u)$ 在 u_0 处连续.

注 2.6.1 条件 (C2.6.1)~(C2.6.7) 是文献中使用的普通条件. 条件 (C2.6.1) 保证了不需要欠光滑 $\hat{g}(\mu)$, 以便可以使用数据驱动方法选择带宽. 条件 (C2.6.2) 对 $K(\cdot)$ 的紧支撑的要求仅是技术上的假定, 在实际运用中通常可以放宽到使用小尾的核; 例如, 标准 Gauss 核. 条件 (C2.6.4) 确保了 $\hat{g}_2$ 和 $\hat{g}$ 的分母以概率 1 有界且下确界大于 0; 由此再结合条件 (C2.6.2) 和 (C2.6.5) 确保了 $\hat{g}_2$ 和 $\hat{g}$ 具有高阶的收敛速度. 条件 (C2.6.6) 保证了 $\hat{\beta}_{\mathrm{EL}}$ 的渐近方差存在. 条件 (C2.6.7) 是基本的假定.

定理 2.6.1 给出了 $\hat{l}(\beta_0)$ 渐近于自由度为 p 的卡方分布.

定理 2.6.1 设条件 (C2.6.1)~(C2.6.5) 成立, 则 $\hat{l}(\beta_0)\xrightarrow{D}\chi_p^2$.

以 $\chi_p^2(1-\alpha)$ 记 χ_p^2 的 α 分位数, $0<\alpha<1$. 使用定理 2.6.1 可以得到 β_0 的近似 $1-\alpha$ 置信域

$$R_\alpha(\beta)=\{\beta|\,\hat{l}(\beta)\leqslant\chi_p^2(1-\alpha)\}.$$

通过极小化 $\hat{l}(\beta)$ 可以得到 β_0 的估计量 $\hat{\beta}_{\mathrm{EL}}$, 称之为极大经验似然估计 (MELE). 不难得到

$$\hat{\beta}_{\mathrm{EL}}=\hat{A}^{-1}n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}w(U_{ij})\widetilde{X}_{ij}\widetilde{Y}_{ij},$$

其中

$$\hat{A}=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}w(U_{ij})\widetilde{X}_{ij}\widetilde{X}_{ij}^{\mathrm{T}}. \tag{2.6.9}$$

估计量 $\hat{\beta}_{\mathrm{EL}}$ 的渐近正态性由下列定理给出.

定理 2.6.2　设条件 (C2.6.1)~(C2.6.6) 成立, 则

$$\sqrt{n}(\hat{\beta}_{\mathrm{EL}}-\beta_0)\xrightarrow{D}N(0,\Sigma),$$

其中 $\Sigma=A^{-1}BA^{-1}$,

$$A=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\{[X_i-g_2(U_i)]^{\mathrm{T}}W_i[X_i-g_2(U_i)]\},$$

$$B=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\big\{[X_i-g_2(U_i)]^{\mathrm{T}}W_i\varepsilon_i\varepsilon_i^{\mathrm{T}}W_i[X_i-g_2(U_i)]\big\}.$$

此处 $X_i=(X_{i1},\cdots,X_{im_i})^{\mathrm{T}}$, $U_i=(U_{i1},\cdots,U_{im_i})^{\mathrm{T}}$, $g_2(U_i)=(g_2(U_{i1}),\cdots,g_2(U_{im_i}))^{\mathrm{T}}$, $\varepsilon_i=(\varepsilon_{i1}\cdots,\varepsilon_{im_i})^{\mathrm{T}}$, $W_i=\mathrm{diag}\big(w(U_{i1}),\cdots,w(U_{im_i})\big)$.

为了应用定理 2.6.2 构造 β_0 的置信域, 我们给出 Σ 的估计量 $\hat{\Sigma}=\hat{A}^{-1}\hat{B}(\hat{\beta}_{\mathrm{EL}})\hat{A}^{-1}$, 这里 $\hat{A}$ 在 (2.6.9) 式中定义, 且

$$\hat{B}(\hat{\beta}_{\mathrm{EL}})=\frac{1}{n}\sum_{i=1}^{n}\hat{Z}_i(\hat{\beta}_{\mathrm{EL}})\hat{Z}_i^{\mathrm{T}}(\hat{\beta}_{\mathrm{EL}}). \tag{2.6.10}$$

可以证明 $\hat{\Sigma}$ 是 Σ 的相合估计. 因而, 由定理 2.6.2 可得

$$\hat{\Sigma}^{-1/2}\sqrt{n}\,(\hat{\beta}_{\mathrm{EL}}-\beta_0)\xrightarrow{D}N(0,I_p), \tag{2.6.11}$$

其中 I_p 是 p 阶单位矩阵. 使用文献 (Arnold, 1981) 的定理 10.2d 得到

$$\big(\hat{\beta}_{\mathrm{EL}}-\beta_0\big)^{\mathrm{T}}n\hat{\Sigma}^{-1}\big(\hat{\beta}_{\mathrm{EL}}-\beta_0\big)\xrightarrow{D}\chi_p^2. \tag{2.6.12}$$

因此, 使用 (2.6.11) 或 (2.6.12) 式可以构造 β_0 的置信区间或置信域.

2. *基准函数的经验似然*

使用估计量 $\hat{\beta}_{\mathrm{EL}}$ 可以在 $[0,1]$ 的内点 u 处构造 $g(u)$ 的估计的经验对数似然比函数, 即

$$\hat{l}(g(u))=-2\max\left\{\sum_{i=1}^{n}\log(np_i)\Big|\,p_i\geqslant 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\hat{\eta}_i(g(u))=0\right\},$$

其中

$$\hat{\eta}_i(g(u)) = \sum_{j=1}^{m_i} \left[Y_{ij} - \hat{\beta}_{\mathrm{EL}}^{\mathrm{T}} X_{ij} - g(u)\right] K\left(\frac{U_{ij}-u}{h}\right).$$

$g(u)$ 的极大经验似然估计可以表示为

$$\hat{g}(u) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} W_{ij}(u)\left[Y_{ij} - \hat{\beta}_{\mathrm{EL}}^{\mathrm{T}} X_{ij}\right] + o_P\left((Nh)^{-1/2}\right), \tag{2.6.13}$$

这里 $W_{ij}(u)$ 在 (2.6.7) 式中定义.

下面考虑在固定点 u_0 处 $\hat{l}(g(u_0))$ 和 $\hat{g}(u_0)$ 的渐近性质. 下文假定 u_0 是 $[0,1]$ 的内点. 由引理 2.6.5 和引理 2.6.6 可得

$$\hat{l}(g(u_0)) = \left(\sum_{i=1}^{n} \hat{\eta}_i(g(u_0))\right)^2 \Big/ \sum_{i=1}^{n} \hat{\eta}_i^2(g(u_0)) + o_P(1). \tag{2.6.14}$$

此外, 由引理 2.6.5 可知 $(Nh)^{-1/2}\sum_{i=1}^{n} \hat{\eta}_i(g(u_0))$ 渐近于正态分布, 其均值为 $b(u_0)$, 方差为 $v^2(u_0)$, 这里

$$b(u_0) = h_0^{5/2}[g'(u_0)f''(u_0) + (1/2)g''(u_0)f(u_0)]\int_{-1}^{1} u^2 K(u)\mathrm{d}u, \tag{2.6.15}$$

$$v^2(u_0) = \sigma_\varepsilon^2(u_0)f(u_0)\int_{-1}^{1} K^2(u)\mathrm{d}u, \tag{2.6.16}$$

h_0 是由条件 (C2.6.1) 定义的常数.

又由引理 2.6.6 可知 $(Nh)^{-1}\sum_{i=1}^{n} \hat{\eta}_i^2(g(u_0))$ 依概率收敛于 $v^2(u_0)$. 这就蕴涵 $\hat{l}(g(u_0))$ 的渐近分布是非中心 χ^2 分布. 为应用 $\hat{l}(g(u_0))$ 构造 $g(u_0)$ 的置信区间, 一个自然的途径是利用 $b(u_0)$ 的估计对经验似然函数进行校正. 注意到

$$(N/h)^{1/2}E\left\{[g(U_{ij}) - g(u_0)]K((U_{ij}-u_0)/h)\right\} = b(u_0) + o(1).$$

因此, 可定义 $b(u_0)$ 的相合估计为

$$\hat{b}(u_0) = \frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}[\hat{g}(U_{ij}) - \hat{g}(u_0)]K\left(\frac{U_{ij}-u_0}{h}\right). \tag{2.6.17}$$

记

$$\xi(g(u)) = \left(2\hat{b}(u)\cdot\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\eta}_i(g(u)) - \hat{b}^2(u)\right)\left(\frac{1}{Nh}\sum_{i=1}^{n}\hat{\eta}_i^2(g(u))\right)^{-1}.$$

那么, $g(u)$ 的校正的经验对数似然比函数可定义为

$$\hat{l}^*(g(u)) = \hat{l}(g(u)) - \xi(g(u)). \tag{2.6.18}$$

我们也可以直接对加权残差 $\hat{l}(g(u))$ 进行调整, 引入辅助随机变量

$$\tilde{\eta}_i(g(u)) = \sum_{j=1}^{m_i} \{Y_{ij} - \hat{\beta}_{\mathrm{EL}}^{\mathrm{T}} X_{ij} - g(u) - [\hat{g}(U_{ij}) - \hat{g}(u)]\} K\left(\frac{U_{ij}-u}{h}\right),$$

因此, $g(u)$ 的残差调整的经验对数似然比函数定义为

$$\tilde{l}(g(u)) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\Big| p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\tilde{\eta}_i(g(u)) = 0\right\}.$$

下面两个定理给出了 $\hat{l}(g(u_0))$, $\hat{l}^*(g(u_0))$, $\tilde{l}(g(u_0))$ 和 $\hat{g}(u_0)$ 的渐近性质.

定理 2.6.3 设条件 (C2.6.1)~(C2.6.7) 成立, 又假设 $K(\cdot)$ 在 $[0,1]$ 上是二次连续可微的. 则当 $g(u_0)$ 是基准函数的真值时, 有 $\tilde{l}(g(u_0)) \xrightarrow{D} \chi_1^2$, $\hat{l}^*(g(u_0)) \xrightarrow{D} \chi_1^2$. 又如果用 $Nh^2/\log N \to \infty$ 和 $Nh^5 \to 0$ 代替条件 (C2.6.1), 则 $\hat{l}(g(u_0)) \xrightarrow{D} \chi_1^2$.

由定理 2.6.3, 可以定义 $g(u_0)$ 的近似 $1-\alpha$ 置信区间

$$\tilde{I}_\alpha(g(u_0)) = \{g(u_0)|\, \tilde{l}(g(u_0)) \leqslant \chi_1^2(\alpha)\},$$

$$\hat{I}_\alpha^*(g(u_0)) = \{g(u_0)|\, \hat{l}^*(g(u_0)) \leqslant \chi_1^2(\alpha)\},$$

$$\hat{I}_\alpha(g(u_0)) = \{g(u_0)|\, \hat{l}(g(u_0)) \leqslant \chi_1^2(\alpha)\}.$$

定理 2.6.4 设条件 (C2.6.1)~(C2.6.7) 成立, 则

$$\sqrt{Nh}\,[\hat{g}(u_0) - g(u_0)] - b(u_0)[f(u_0)]^{-1} \xrightarrow{D} N(0, \sigma^2(u_0)),$$

其中 $\sigma^2(u_0) = v^2(u_0)[f(u_0)]^{-2}$, $b(u_0)$ 和 $v^2(u_0)$ 分别在 (2.6.15) 和 (2.6.16) 式中定义. 又如果用 $Nh^2/\log N \to \infty$ 和 $Nh^5 \to 0$ 代替条件 (C2.6.1), 则

$$\sqrt{Nh}\,[\hat{g}(u_0) - g(u_0)] \xrightarrow{D} N(0, \sigma^2(u_0)).$$

现在构造 $\sigma^2(u_0)$ 的估计量. 因 $\sigma^2(u_0)$ 依赖于 $f(u_0)$ 和 $v^2(u_0)$, 首先需要估计 $f(u_0)$ 和 $v^2(u_0)$. 使用核光滑可以估计密度 $f(u_0)$. 取一个核函数 $K(\cdot)$ 和带宽 h, 假定 $K(\cdot)$ 满足条件 (C2.6.2). 如果 h 满足 $h \to 0$ 和 $Nh^2/\log N \to \infty$, 那么 $f(u_0)$ 和 $v^2(u_0)$ 的相合估计可分别定义为

$$\hat{f}(u_0) = \frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K\left(\frac{U_{ij}-u_0}{h}\right),$$

$$\hat{v}^2(u_0) = \frac{1}{Nh}\sum_{i=1}^{n}\hat{\eta}_i^2(\hat{g}(u_0)). \tag{2.6.19}$$

因此, 用 $\hat{f}(u_0)$ 和 $\hat{v}^2(u_0)$ 分别代替 $\sigma^2(u_0)$ 中的 $f(u_0)$ 和 $v^2(u_0)$, 可以得到 $\sigma^2(u_0)$ 的相合估计 $\hat{\sigma}^2(u_0)$. 此外, 由引理 2.6.7 知 $\hat{b}(u_0)$ 是 $b(u_0)$ 的相合估计. 于是, 由定理 2.6.4 可得

$$\hat{\sigma}^{-1}(u_0)\{\sqrt{Nh}\,[\hat{g}(u_0)-g(u_0)]-\hat{b}(u_0)(\hat{f}(u_0))^{-1}\} \xrightarrow{D} N(0,1). \tag{2.6.20}$$

使用 (2.6.20) 式可以得到 $g(u_0)$ 的近似 $1-\alpha$ 置信区间

$$\hat{g}(u_0)-(Nh)^{-1/2}\hat{b}(u_0)(\hat{f}(u_0))^{-1} \pm z_{1-\alpha/2}(Nh)^{-1/2}\hat{\sigma}(u_0),$$

其中 $z_{1-\alpha/2}$ 是标准正态分布的 $1-\alpha/2$ 分位点.

3. 带宽选择

因为个体之间是独立的, Rice 和 Silverman(1991) 使用 "去一个体" 交叉验证法选择带宽, 其准则为

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[Y_{ij}-X_{ij}^{\mathrm{T}}\hat{\beta}^{(-i)}-\hat{g}_h^{(-i)}(U_{ij})\right]^2,$$

其中 $\hat{\beta}^{(-i)}$ 和 $\hat{g}_h^{(-i)}(U_{ij})$ 分别是 β_0 和 $g(U_{ij})$ 的去掉第 i 个个体的观察得到的估计量. 一个交叉验证带宽定义为 $h_{\mathrm{CV}} = \inf_{h>0}\mathrm{CV}(h)$. Wang(2003) 和 Hu 等 (2004) 的渐近偏差和方差结构可以用来证明 h_{CV} 的阶为 $n^{-1/5}$.

4. 模拟和应用

模拟例子 为实施模拟, 我们产生 500 个数据集, 每个数据集包含 $n=100$ 个个体且每个个体具有 $m_i=3$ 次观测. 协变量 X_{ij} 设置为 $X_{ij}=(X_{1ij},X_{2i})^{\mathrm{T}}$, 其中 X_{1ij} 是随时间变化的协变量; X_{2i} 由成功概率为 0.5 的 Bernoulli 随机变量产生. 我们按照 $X_{1ij}=\xi_i+e_{ij}$ 和 $U_{ij}=\xi_i+e'_{ij}$ 产生 X_{1ij} 和 U_{ij}, 这里 ξ_i 服从 $(-1,1)$ 上的均匀分布, 而 e_{ij} 和 e'_{ij} 相互独立且都服从 $(-1,1)$ 上的均匀分布. 这种设置允许 X_{1ij} 与 U_{ij} 在它们的重复测量之间相关且具有相关系数 0.5. Y_{ij} 由模型 (2.6.1) 产生, 其中 $\beta_1=1$, $\beta_2=2$, $g(u)=2\sin(\pi u)$, 误差 ε_{ij} 服从具有方差为 1、内相关系数为 0.8 的正态分布.

对每个模拟的数据集, 我们计算 β_0 和 $g(u)$ 的经验似然比统计量和估计值. 核函数取为 Epanechnikov 核 $K(u)=0.75(1-u^2)_+$, 权函数 $w(t)$ 取为区间 $[-0.99,0.99]$ 上的示性函数, 采用 "去一个体" 交叉验证法选择带宽 h_{CV}. $\hat{f}(u_0)$ 的带宽也利用交叉验证法选择. 为简单起见, 由 (2.6.17) 和 (2.6.19) 式定义的估计量 $\hat{b}(u_0)$ 和 $\hat{v}^2(u_0)$

的带宽也取为 h_{CV}. 用经验似然 (EL) 和正态逼近 (NA) 两种方法比较 β_0 的置信域的面积和覆盖水平. 模拟结果展示在图 2.6.1 中.

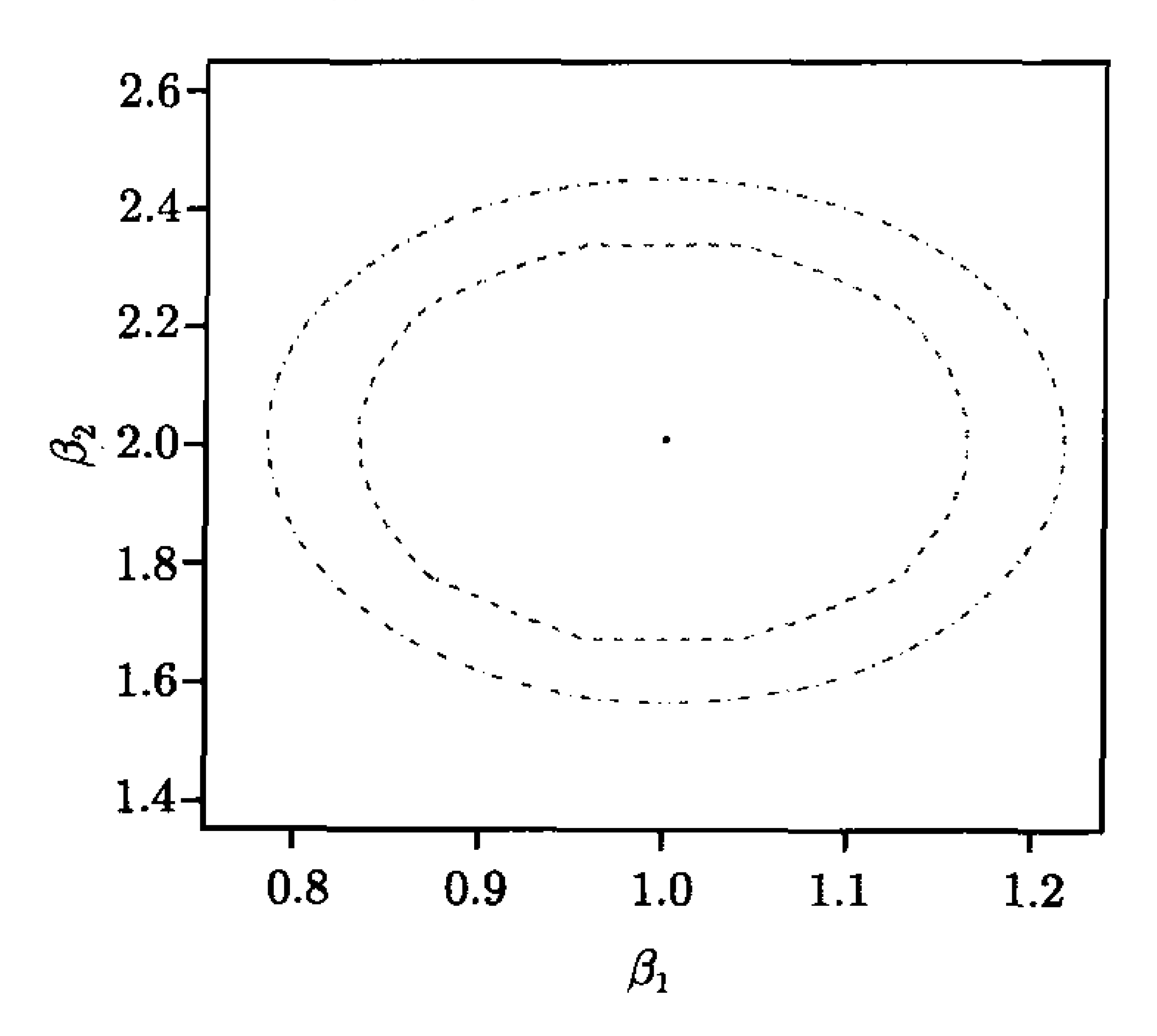


图 2.6.1　基于 EL(虚线) 和 NA(点虚线), 参数 (β_1, β_2) 的近似 0.95 置信域

图 2.6.1 说明 EL 比 NA 给出更小的置信域, 且 EL 的覆盖概率是 0.936, 而 NA 的覆盖概率是 0.934. 这就说明了 EL 比 NA 给出的置信域有更高的精度.

对于函数 $g(u)$, 用三种方法作比较: 残差调整的经验似然 (RAEL), 均值校正的经验似然 (MCEL) 和正态逼近 (NA). 我们没有使用估计的经验似然比, 因为它在计算中需要对基准函数的估计做欠光滑, 这就为带宽的选择带来困难. 我们从置信区间的长度和覆盖概率两个方面进行比较. 模拟结果展示在图 2.6.2 中.

从图 2.6.2 可以看出, RAEL 比其他两个方法给出更短的置信区间且有稍高的覆盖概率. 这说明残差调整的经验似然基本达到了满意的效果.

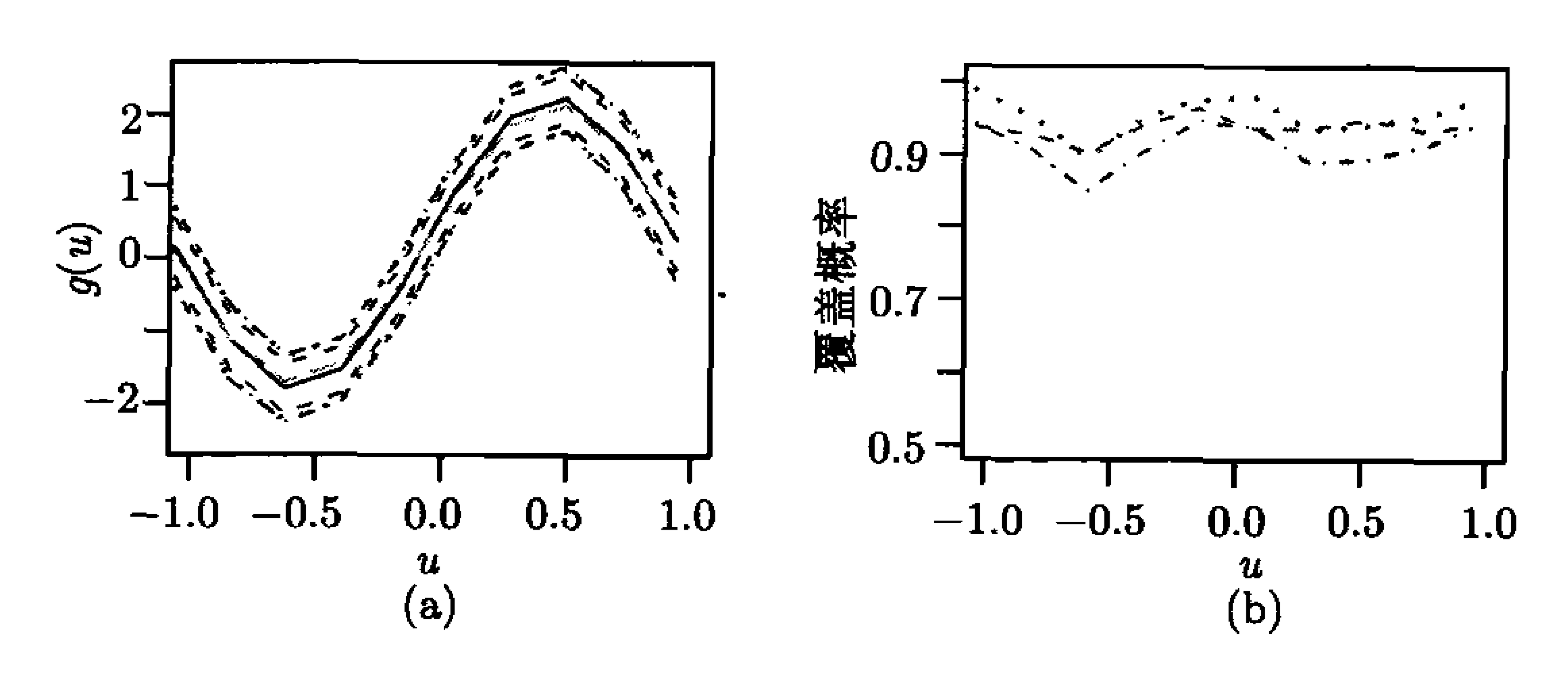


图 2.6.2　基于 RAEL(虚线)、MCEL(点线) 和 NA(点虚线), $g(u)$ 的近似 0.95 置信区间及其覆盖概率. (a) 逐点置信区间、真曲线 (实线) 和估计的曲线 (长虚线). (b) 覆盖概率

一个实例.　我们把经验似然方法应用于纵向荷尔蒙研究. 收集 34 个健康妇女在一个月经周期的尿样, 每隔一天试验尿的孕激素. 在 34 个参与者中, 每个妇女

按时提供 11 到 28 次观测, 共得到 492 个观测值. 平均每个妇女做 14.5 次观测. 这些妇女的月经周期的长度分布于 23 到 56 天, 平均为 29.6 天. 为方便起见, 每个妇女的月经周期长度一律统一为 28 天 (参考周期). Zhang 等 (1998) 使用半参数混合模型拟合该数据集, 其响应变量取为孕激素值的对数, 参数部分的协变量分别是年龄 (AEG) 和体重指标 (BMI), 变量 U 代表时间 (天). He 等 (2002) 使用下列半参数模型:

$$Y_{ij} = \beta_1 \mathrm{AGE}_i + \beta_2 \mathrm{BMI}_i + g(U_{ij}) + \varepsilon_{ij}$$

拟合该数据集, 并得到了 β_1 和 β_2 的估计. 我们的目的是给出 (β_1, β_2) 的置信域和 $g(u)$ 的逐点置信区间. 仍使用 Epanechnikov 核和带宽 h_{CV}. 权函数取为 $w(t) = 1$. 模拟结果展示在图 2.6.3 中. 从图 2.6.3 可以看出, EL 比 NA 的效果好.

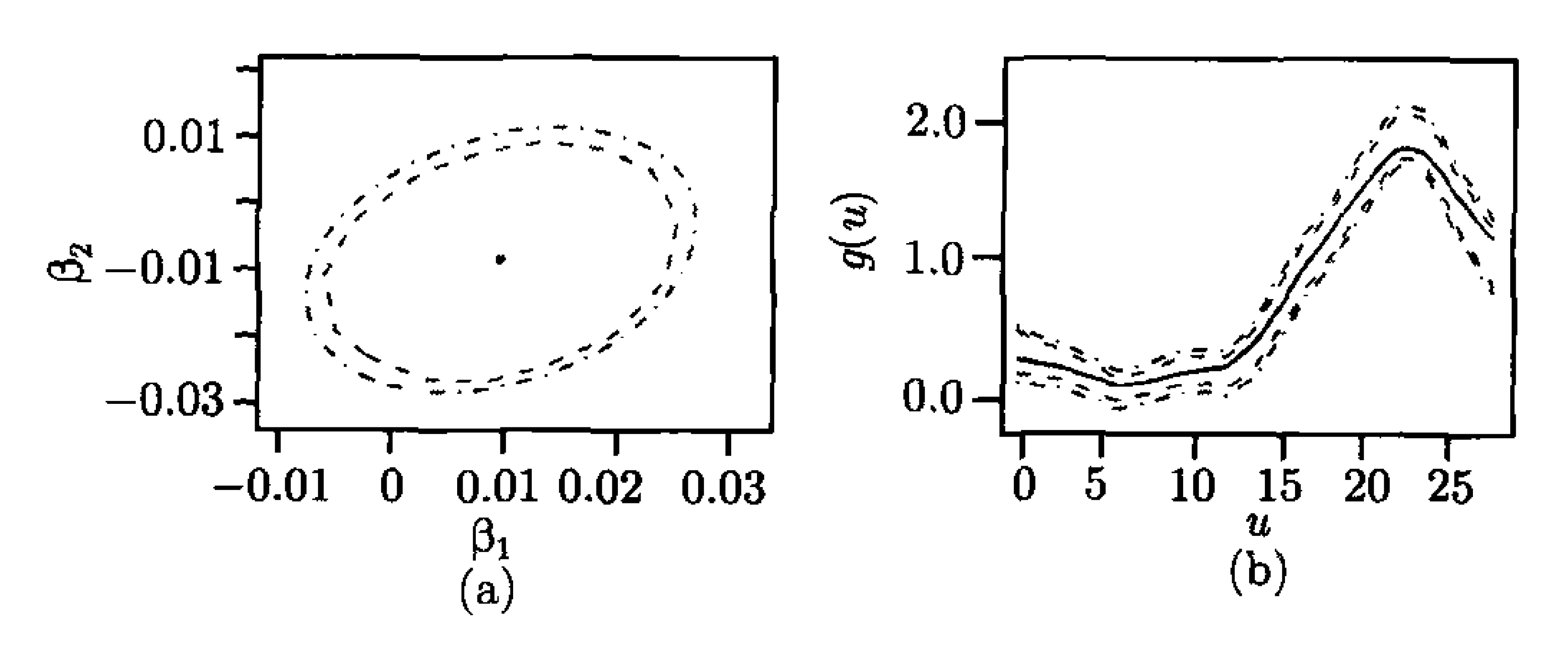


图 2.6.3 应用于纵向荷尔蒙研究. 基于 EL 或 RAEL(虚线) 和 NA(点虚线), (β_1, β_2) 和 $g(u)$ 的近似 0.95 置信域和逐点置信区间. (a) 是对年龄和体重指标, (b) 是对 $g(u)$; 实曲线是 $g(u)$ 的估计曲线

5. **定理的证明**

为简便起见, 下文用 c 表示正的常数, 每次出现时可以代表不同的值. 在证明本节的主要结果之前, 首先给出下面几个引理.

引理 2.6.1 设条件 (C2.6.1)~(C2.6.5) 成立, 则

$$\sup_{a\leqslant u\leqslant b} E\big[\|g_2(U_{ij}) - \hat{g}_2(U_{ij})\|^2 | U_{ij} = u\big] = O\big(n^{-1}h^{-1} + h^4\big),$$

$$\sup_{a\leqslant u\leqslant b} E\big[|g(U_{ij}) - \hat{g}_*(U_{ij})|^2 | U_{ij} = u\big] = O\big(n^{-1}h^{-1} + h^4\big),$$

其中 $\hat{g}_*(u) = \hat{g}_1(u) - \beta_0^{\mathrm{T}} \hat{g}_2(u)$.

证 仅证明第一个等式, 第二个等式可用同样的方法证明. 用 $g_{2r}(u)$, $\hat{g}_{2r}(u)$

和 X_{klr} 分别表示 $g_2(u)$, $\hat{g}_2(u)$ 和 X_{kl} 的第 r 个分量. 由 Cauchy-Schwarz 不等式得

$$\begin{aligned}
&E\big[|g_{2r}(U_{ij})-\hat{g}_{2r}(U_{ij})|^2|U_{ij}=u\big]\\
&\leqslant 2E\left[\Big|g_{2r}(U_{ij})-\sum_{k=1}^{n}\sum_{l=1}^{m_k}W_{kl}(U_{ij})g_{2r}(U_{kl})\Big|^2\Big|U_{ij}=u\right]\\
&\quad+2E\left[\Big|\sum_{k=1}^{n}\sum_{l=1}^{m_k}W_{kl}(U_{ij})(g_{2r}(U_{kl})-X_{klr})\Big|^2\Big|U_{ij}=u\right]\\
&\equiv I_1(u)+I_2(u).
\end{aligned}\tag{2.6.21}$$

首先证明

$$\sup_{a\leqslant u\leqslant b}I_1(u)=O\big(n^{-1}h+h^4\big).\tag{2.6.22}$$

使用密度函数核估计的常规证明方法可以证得

$$\sup_{a\leqslant u\leqslant b}|\hat{f}(u)-f(u)|\longrightarrow 0,\quad \text{a.s.}\tag{2.6.23}$$

因此, 当 n 充分大时, 以概率 1 有 $\sup\limits_{a\leqslant u\leqslant b}|\hat{f}(u)-f(u)|\leqslant c_1/2$. 由此再利用条件 (C2.6.4) 可得: 当 n 充分大时,

$$\inf_{a\leqslant u\leqslant b}|\hat{f}(u)|\geqslant\inf_{a\leqslant u\leqslant b}f(u)-\sup_{a\leqslant u\leqslant b}|\hat{f}(u)-f(u)|\geqslant c_1/2,\quad\text{a.s.}\tag{2.6.24}$$

记 $\psi_{kl}(U_{ij})=K((U_{kl}-U_{ij})/h)[g_{2r}(U_{ij})-g_{2r}(U_{kl})]$. 使用条件 (C2.6.2)、(C2.6.4) 和 (C2.6.5) 可得

$$\begin{aligned}
&\sup_{a\leqslant u\leqslant b}\big|E[\psi_{kl}(U_{ij})|U_{ij}=u]\big|=O\big(h^3\big),\\
&\sup_{a\leqslant u\leqslant b}E[\psi_{kl}^2(U_{ij})|U_{ij}=u]=O\big(h^3\big).
\end{aligned}$$

结合 (2.6.24) 式可证得

$$\begin{aligned}
I_1(u)&\leqslant c(nh)^{-2}E\left[\Big|\sum_{k=1}^{n}\sum_{l=1}^{m_k}\big\{\psi_{kl}(U_{ij})-E[\psi_{kl}(U_{ij})|U_{ij}]\big\}\Big|^2\Big|U_{ij}=u\right]\\
&\quad+c(nh)^{-2}\left\{\sum_{k=1}^{n}\sum_{l=1}^{m_k}E\big[\psi_{kl}(U_{ij})|U_{ij}=u\big]\right\}^2\\
&\leqslant cn^{-1}h+ch^4,\quad\forall\, u\in[a,b].
\end{aligned}$$

这就证明了 (2.6.22) 式. 由 (2.6.24) 式和条件 (C2.6.2)~(C2.6.4) 可得

$$\begin{aligned}
I_2(u)&=2\sum_{k=1}^{n}E\left[\sum_{l=1}^{m_k}W_{kl}(U_{ij})(g_{2r}(U_{kl})-X_{klr})\Big|U_{ij}=u\right]^2\\
&\leqslant c(nh)^{-1},\quad\forall\, u\in[a,b].
\end{aligned}$$

上式与 (2.6.21) 和 (2.6.22) 式结合即可得到引理 2.6.1 的第一个等式.

引理 2.6.2 设条件 (C2.6.1)~(C2.6.5) 成立, 则

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0)\xrightarrow{D}N(0,B),$$

其中 B 在定理 2.6.2 中定义.

证 记 $\breve{\mu}(u)=g_2(u)-\hat{g}_2(u)$, $\breve{g}_*(u)=g(u)-\hat{g}_*(u)$. 简单计算得

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0)=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}w(U_{ij})\breve{X}_{ij}\varepsilon_{ij}\\&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}w(U_{ij})\breve{\mu}(U_{ij})\varepsilon_{ij}\\&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}w(U_{ij})\breve{X}_{ij}\breve{g}_*(U_{ij})\\&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}w(U_{ij})\breve{\mu}(U_{ij})\breve{g}_*(U_{ij})\\\equiv&J_1+J_2+J_3+J_4.\end{aligned}$$

首先处理 J_1. 容易验证 $E(J_1)=0$ 和 $\mathrm{var}(J_1)=B+o(1)$. 利用引理 2.6.2 的条件可以推出 J_1 满足 Cramer-Wold 定理的条件和 Lindeberg 条件 (Serfling, 1980). 于是, $J_1\xrightarrow{D}N(0,B)$. 因此, 为证引理 2.6.2, 仅需证明 $J_\nu\xrightarrow{P}0$, $\nu=2,3,4$. 现在处理 J_2. 由引理 2.6.1 得

$$\begin{aligned}E\big(||J_2||^2\big)&\leqslant cn^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}E\{w(U_{ij})E[||\breve{\mu}(U_{ij})||^2|U_{ij}]\}\\&\leqslant c[(nh)^{-1}+h^4]\longrightarrow 0.\end{aligned}$$

因此, $J_2\xrightarrow{P}0$. 类似 (2.6.22) 式的证明可以证得

$$\sup_{a\leqslant u\leqslant b}E\left[\left(g(U_{ij})-\sum_{k=1}^{n}\sum_{l=1}^{m_k}W_{kl}(U_{ij})g(U_{kl})\right)^2\middle|U_{ij}=u\right]=O\big(n^{-1}h+h^4\big).\quad(2.6.25)$$

用 $J_{3,s}$ 和 $\breve{X}_{ijs}(u)$ 分别记 J_3 和 $\breve{X}_{ij}(u)$ 的第 s 个分量. 由 (2.6.25) 式和条件

(C2.6.2)~(C2.6.5) 得

$$\begin{aligned}E(J_{3,s}^2) &\leqslant 2n^{-1}E\left[\sum_{k=1}^{n}\sum_{l=1}^{m_k}\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i}w(U_{ij})W_{kl}(U_{ij})\check{X}_{ijs}\right)\varepsilon_{kl}\right]^2\\ &\quad+2n^{-1}E\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}w(U_{ij})\check{X}_{ijs}\left(g(U_{ij})-\sum_{k=1}^{n}\sum_{l=1}^{m_k}W_{kl}(U_{ij})g(U_{kl})\right)\right]^2\\ &\leqslant c(nh)^{-1}+ch^4\longrightarrow 0.\end{aligned}$$

这就得到 $J_3 \xrightarrow{P} 0$. 利用引理 2.6.1 和 Cauchy-Schwarz 不等式, 有

$$\begin{aligned}E(\|J_4\|) &\leqslant \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}E\Big\{w(U_{ij})\big[E\big(\|\check{\mu}(U_{ij})\|^2|U_{ij}\big)\big]^{1/2}\big[E\big(\check{g}_*^2(U_{ij})|U_{ij}\big)\big]^{1/2}\Big\}\\ &\leqslant c(nh^2)^{-1/2}+c\sqrt{n}h^4\longrightarrow 0.\end{aligned}$$

由此推出 $J_4 \xrightarrow{P} 0$. 这就完成了引理 2.6.2 的证明.

引理 2.6.3　设条件 (C2.6.1)~(C2.6.5) 成立, 则

$$\hat{B}(\beta_0)=\frac{1}{n}\sum_{i=1}^{n}\hat{Z}_i(\beta_0)\hat{Z}_i^{\mathrm{T}}(\beta_0)\xrightarrow{P} B.$$

证　仍使用引理 2.6.2 的证明中的记号, 并记 $U_i=\sum_{j=1}^{m_i}w(U_{ij})\check{X}_{ij}\varepsilon_{ij}$,

$$R_i=\sum_{j=1}^{m_i}w(U_{ij})\big\{\check{\mu}(U_{ij})\varepsilon_{ij}+[\check{X}_{ij}+\check{\mu}(U_{ij})]\check{g}(U_{ij})\big\}.$$

那么有 $\hat{Z}_i(\beta_0)=U_i+R_i$,

$$\begin{aligned}\hat{B}(\beta_0)&=\frac{1}{n}\sum_{i=1}^{n}U_iU_i^{\mathrm{T}}+\frac{1}{n}\sum_{i=1}^{n}R_iR_i^{\mathrm{T}}+\frac{1}{n}\sum_{i=1}^{n}U_iR_i^{\mathrm{T}}+\frac{1}{n}\sum_{i=1}^{n}R_iU_i^{\mathrm{T}}\\ &\equiv D_1+D_2+D_3+D_4.\end{aligned}$$

利用大数定律可以得到 $D_1 \xrightarrow{P} B$. 因而, 如果证明了 $D_\nu \xrightarrow{P} 0$, $\nu=2,3,4$, 即可证得引理 2.6.3. 先考虑 D_2, 以 $D_{2,rs}$ 记 D_2 的 (r,s) 元素, R_{ir} 记 R_i 的第 r 个分量. 使用 Cauchy-Schwarz 不等式可得

$$|D_{2,rs}|\leqslant\left(\frac{1}{n}\sum_{i=1}^{n}R_{ir}^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}R_{is}^2\right)^{1/2}. \tag{2.6.26}$$

利用引理 2.6.1 可以推出 $n^{-1}\sum_{i=1}^{n} R_{ir}^2 \xrightarrow{P} 0$. 由此再结合 (2.6.26) 式就证明了 $D_2 \xrightarrow{P} 0$. 类似地, 可以证明 $D_\nu \xrightarrow{P} 0$, $\nu = 3, 4$. 这就完成了引理 2.6.3 的证明.

引理 2.6.4 设条件 (C2.6.1)~(C2.6.5) 成立, 则

$$\max_{1\leqslant i\leqslant n} ||\hat{Z}_i(\beta_0)|| = o_P\big(n^{1/2}\big), \tag{2.6.27}$$

$$||\lambda|| = O_P\big(n^{-1/2}\big). \tag{2.6.28}$$

证 利用引理 2.6.3 并类似于 Owen(1990) 的证明方法可以证得 (2.6.28) 式. 现在证明 (2.6.27) 式. 仍使用引理 2.6.2 的证明中的记号. 显然

$$\begin{aligned}\max_{1\leqslant i\leqslant n} ||\hat{Z}_i(\beta_0)|| \leqslant& \max_{1\leqslant i\leqslant n}\left|\left|\sum_{j=1}^{m_i} w(U_{ij})\check{X}_{ij}\varepsilon_{ij}\right|\right| + \max_{1\leqslant i\leqslant n}\left|\left|\sum_{j=1}^{m_i} w(U_{ij})\check{\mu}(U_{ij})\varepsilon_{ij}\right|\right| \\ &+ \max_{1\leqslant i\leqslant n}\left|\left|\sum_{j=1}^{m_i} w(U_{ij})\big[\check{X}_{ij} + \check{\mu}(U_{ij})\big]\check{g}_*(U_{ij})\right|\right| \\ \equiv& M_1 + M_2 + M_3.\end{aligned}$$

利用文献 (Owen, 1990) 的引理 2.6.3 可得 $M_1 = o_P\big(n^{1/2}\big)$. 使用 Cauchy-Schwarz 不等式和引理 2.6.1 可以推出 $M_\nu = o_P\big(n^{1/2}\big)$, $\nu = 2, 3$. 这就证明了引理 2.6.4.

引理 2.6.5 设条件 (C2.6.1)~(C2.6.7) 成立. 如果 $g(u_0)$ 是基准函数的真值, 则

$$\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\eta}_i(g(u_0)) - b(u_0) \xrightarrow{D} N(0, v^2(u_0)).$$

其中 $v^2(u_0)$ 在 (2.6.16) 式中定义.

证 注意到 $\tilde{l}(g(u_0))$ 中定义的记号 $\hat{\eta}_i(g(u_0))$, 有

$$\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\eta}_i(g(u_0)) - b(u_0) = S_1(u_0) + S_2(u_0) + S_3(u_0), \tag{2.6.29}$$

其中

$$S_1(u_0) = \frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K\left(\frac{U_{ij}-u_0}{h}\right)\varepsilon_{ij},$$

$$S_2(u_0) = \frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{K\left(\frac{U_{ij}-u_0}{h}\right)[g(U_{ij}) - g(u_0)] - (h/N)^{1/2}b(u_0)\right\}$$

$$S_3(u_0) = \frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K\left(\frac{U_{ij}-u_0}{h}\right)X_{ij}^{\mathrm{T}}(\beta_0 - \hat{\beta}_{\mathrm{EL}}).$$

不难证明: $E\big[S_1(u_0)\big] = 0$, $\text{var}\big(S_1(u_0)\big) = \upsilon^2(u_0) + o(1)$. 可以验证 $S_1(u_0)$ 满足 Cramer-Wold 定理的条件和 Lindeberg 条件 (Serfling, 1980). 因此,

$$S_1(u_0) \xrightarrow{D} N(0, \upsilon^2(u_0)). \tag{2.6.30}$$

可以证明 $\text{var}(S_2(u_0)) = o(1)$. 因而,

$$S_2(u_0) \xrightarrow{P} 0. \tag{2.6.31}$$

容易证得

$$\frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K\left(\frac{U_{ij}-u_0}{h}\right) X_{ij} = O_P(1).$$

由定理 2.6.2 知 $\hat{\beta}_{\text{EL}} - \beta_0 = O_P(N^{-1/2})$. 因此得到 $S_3(u_0) = O_P\big(h^{1/2}\big)$. 最后, 结合 (2.6.29)~(2.6.31) 式即可完成引理 2.6.5 的证明.

引理 2.6.6　设定理 2.6.3 的条件成立. 如果 $g(u_0)$ 是基准函数的真值, 则

$$\frac{1}{Nh}\sum_{i=1}^{n}\hat{\eta}_i^2(g(u_0)) \xrightarrow{P} \upsilon^2(u_0),$$
$$\max_{1\leqslant i\leqslant n} \|\hat{\eta}_i(g(u_0))\| = o_P\big((Nh)^{1/2}\big),$$

这里 $\upsilon^2(u_0)$ 由 (2.6.16) 式定义.

类似引理 2.6.3 和引理 2.6.4 的证明方法可以证明引理 2.6.6, 这里省略其证明.

引理 2.6.7　设定理 2.6.3 的条件成立, 则 $\hat{b}(u_0) \xrightarrow{P} b(u_0)$.

证　记 $\varphi_{ij}(u_0) = \big[g(U_{ij}) - g(u_0)\big]K((U_{ij}-u_0)/h)$, $\hat{\varphi}_{ij}(u_0) = \big[\hat{g}(U_{ij}) - \hat{g}(u_0)\big]$ $\cdot K((U_{ij}-u_0)/h)$. 则有

$$\begin{aligned}\hat{b}(u_0) - b(u_0) &= \frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\hat{\varphi}_{ij}(u_0) - \varphi_{ij}(u_0)\right] \\ &\quad + \frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\varphi_{ij}(u_0) - (h/N)^{1/2}b(u_0)\right] \\ &= H_1(u_0) + H_2(u_0).\end{aligned}$$

由条件 (C2.6.1)~(C2.6.5) 和 Taylor 公式可得

$$\begin{aligned}\hat{\varphi}_{ij}(u_0) - \varphi_{ij}(u_0) = \big\{&[\hat{g}'(u_0) - g'(u_0)](U_{ij}-u_0) \\ &+ o_P(|U_{ij}-u_0|^2)\big\}K((U_{ij}-u_0)/h).\end{aligned}$$

使用条件 (C2.6.1) 和 (C2.6.2) 可以证明

$$\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(U_{ij}-u_0)^l K\left(\frac{U_{ij}-u}{h}\right) = O_P(1), \quad l = 1, 2$$

和 $\hat{g}'(u_0)-g'(u_0)\xrightarrow{P}0$. 因此, $H_1(u_0)\xrightarrow{P}0$. 容易证明 $H_2(u_0)\xrightarrow{P}0$. 这就完成了引理 2.6.7 的证明.

定理 2.6.1 的证 利用 Lagrange 乘子法, 可以把 $\hat{l}(\beta_0)$ 表示为

$$\hat{l}(\beta_0)=2\sum_{i=1}^{n}\log\big(1+\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0)\big), \tag{2.6.32}$$

其中 $\lambda=\lambda(\beta_0)$ 是 $p\times 1$ 向量且满足

$$\sum_{i=1}^{n}\frac{\hat{Z}_i(\beta_0)}{1+\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0)}=0. \tag{2.6.33}$$

将 (2.6.32) 式进行 Taylor 展开, 并利用引理 2.6.2~ 引理 2.6.4 可得

$$\hat{l}(\beta_0)=2\sum_{i=1}^{n}\left\{\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0)-\left[\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0)\right]^2/2\right\}+o_P(1). \tag{2.6.34}$$

利用 (2.6.33) 式可推得

$$\begin{aligned}0&=\sum_{i=1}^{n}\frac{\hat{Z}_i(\beta_0)}{1+\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0)}\\&=\sum_{i=1}^{n}\hat{Z}_i(\beta_0)-\sum_{i=1}^{n}\hat{Z}_i(\beta_0)\hat{Z}_i^{\mathrm{T}}(\beta_0)\lambda+\sum_{i=1}^{n}\frac{\hat{Z}_i(\beta_0)\left[\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0)\right]^2}{1+\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0)}.\end{aligned}$$

由上式再利用引理 2.6.2~ 引理 2.6.4 可以证得

$$\sum_{i=1}^{n}[\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0)]^2=\sum_{i=1}^{n}\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0)+o_P(1),$$

$$\lambda=\left[\sum_{i=1}^{n}\hat{Z}_i(\beta_0)\hat{Z}_i^{\mathrm{T}}(\beta_0)\right]^{-1}\sum_{i=1}^{n}\hat{Z}_i(\beta_0)+o_P\big(n^{-1/2}\big).$$

因此, 由 (2.6.34) 式得

$$\hat{l}(\beta_0)=\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0)\right]^{\mathrm{T}}\hat{B}(\beta_0)^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0)\right]+o_P(1). \tag{2.6.35}$$

这里 $\hat{B}(\beta_0)$ 由 (2.6.10) 式定义. 利用 (2.6.35)、引理 2.6.2 和引理 2.6.3 即可完成定理 2.6.1 的证明.

定理 2.6.2 的证 经过计算可得

$$\hat{\beta}_{\mathrm{EL}}-\beta_0=\hat{A}^{-1}n^{-1}\sum_{i=1}^{n}\hat{Z}_i(\beta_0). \tag{2.6.36}$$

利用引理 2.6.1 可以证明

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} w(U_{ij})\widetilde{X}_{ij}\widetilde{X}_{ij}^{\mathrm{T}} \xrightarrow{P} A. \tag{2.6.37}$$

最后, 运用 (2.6.36) 和 (2.6.37) 式以及引理 2.6.2 和 Slutsky 定理即可完成定理 2.6.2 的证明.

定理 2.6.3 的证 采用与定理 2.6.1 类似的论证方法可证得 (2.6.14) 式. 记 $\hat{\gamma}(g(u_0)) = (Nh)^{-1}\sum_{i=1}^{n}\hat{\eta}_i^2(g(u_0))$. 由 (2.6.14) 和 (2.6.18) 式得

$$\begin{aligned}\tilde{l}^*(g(u_0)) &= [\hat{\gamma}(g(u_0))]^{-1}\left(\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\eta}_i(g(u_0)) - b(u_0)\right)^2 \\ &\quad + 2[\hat{\gamma}(g(u_0))]^{-1}\left[b(u_0) - \hat{b}(u_0)\right]\left(\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\eta}_i(g(u_0))\right) \\ &\quad + [\hat{\gamma}(g(u_0))]^{-1}\left[\hat{b}^2(u_0) - b^2(u_0)\right] + o_P(1) \\ &\equiv M_1(u_0) + M_2(u_0) + M_3(u_0) + o_P(1).\end{aligned}$$

利用引理 2.6.5~ 引理 2.6.7 可以得到 $M_1(u_0) \xrightarrow{D} \chi_1^2$, $M_\nu(u_0) \xrightarrow{P} 0$, $\nu = 2, 3$. 这就证明了 $\tilde{l}^*(g(u_0)) \xrightarrow{D} \chi_1^2$.

在条件 $Nh^2/\log n \to \infty$ 和 $Nh^5 \to 0$ 下, 类似引理 2.6.5 的证法可以证明:

$$\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\eta}_i(g(u_0)) \xrightarrow{D} N(0, v^2(u_0)). \tag{2.6.38}$$

利用 (2.6.14)、(2.6.38) 和引理 2.6.6 就证明了 $\tilde{l}(g(u_0)) \xrightarrow{D} \chi_1^2$.

定理 2.6.4 的证 类似 Qin 和 Lawless(1994) 的证明方法可以证得

$$\sqrt{Nh}\big(\hat{g}(u_0) - g(u_0)\big) = \frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\eta}_i(g(u_0))/\hat{f}(u_0) + o_P(1).$$

再利用 (2.6.23) 式和引理 2.6.5 即可完成定理 2.6.4 的证明.

2.6.6 计数过程方法

假设有 n 个个体的随机样本. 对第 i 个个体, 在时间点 t 处采集响应变量 $Y_i(t)$ 和协变量 $X_i(t)$, 其中 $t = t_{i1}, \cdots, t_{im_i}$, $i = 1, \cdots, n$, m_i 为在第 i 个个体上观察的总数. 那么, 部分线性模型可表示为

$$Y_i(t_{ij}) = \beta_0^{\mathrm{T}} X_i(t_{ij}) + \theta(t_{ij}) + \varepsilon_i(t_{ij}), \quad i = 1, \cdots, n,\ j = 1, \cdots, m_i, \tag{2.6.39}$$

其中 β_0 为 p 维未知参数向量, $\theta(t)$ 为 $[0,1]$ 上的未知基准函数. 假设来自不同个体的 $(X_i(t_{ij}),Y_i(t_{ij}))$ 和 $\varepsilon_i(t_{ij})$ 独立, 且 $E[\varepsilon_i(t_{ij})|X_i(t_{ij})]=0$. 又设 m_i 有界但 n 趋于无穷大.

Lin 和 Ying(2001) 把计数过程方法引入到估计方案中来. 基本思想如下. 用计数过程

$$N_i(t)\equiv\sum_{j=1}^{m_i}I(t_{ij}\leqslant t)$$

描述在第 i 个个体上观察的时间点, 其中 $I(\cdot)$ 为示性函数. 我们把观察时间看作结束跟踪时删失的任意计数过程的实现. 即 $N_i(t)=N_i^*(t\wedge C_i)$, 其中 $N_i^*(t)$ 为连续时间 $t\in[0,1]$ 的计数过程, C_i 为跟踪或删失时间, $t\wedge C_i=\min\{t,C_i\}$. 遵照 (Lin and Ying, 2001), 假定计数过程 $N_i(t)$ 为来自某个总体的随机样本. $X_i(t)$ 和 $Y_i(t)$ 都在跳跃点 $N_i(t)$ 处被观察. 允许删失时间 C_i 以任意方式依赖于协变量 $X_i(t)$. 本节假定删失机制是无信息的, 即

$$E[Y_i(t)|X_i(t),C_i\geqslant t]=E[Y_i(t)|X_i(t)],\quad i=1,\cdots,n.$$

使用上述计数过程, 定义加权最小二乘函数

$$\sum_{i=1}^{n}\int_0^{\infty}w(t)[Y_i(t)-\beta^{\mathrm{T}}X_i(t)-\theta(t)]^2\mathrm{d}N_i(t),\tag{2.6.40}$$

其中 $w(t)$ 是可能依赖于数据的加权函数.

Lin 和 Ying(2001) 的方法的实质是: 首先估计函数 $\theta(t)$, 然后应用替换技术估计 β_0. Lin 和 Ying(2001) 考虑了两种情况, 即依赖于潜在观察时间是否独立于协变量 $X(t)$. 当潜在观察时间依赖于协变量时, 他们假定

$$E[\mathrm{d}N_i^*(t)|X_i(t),Y_i(t),C_i\geqslant t]=\exp\{\gamma_0^{\mathrm{T}}X_i(t)\}\mathrm{d}\Lambda(t),\quad i=1,\cdots,n,$$

其中 γ_0 是未知的参数向量, $\Lambda(\cdot)$ 是任意不减函数. 当 $\gamma_0=0$ 时, 观察时间独立于协变量. Lin 和 Ying(2001) 使用

$$\hat{\theta}(t;\beta,\gamma)=\bar{Y}(t,\gamma)-\beta^{\mathrm{T}}\bar{X}(t,\gamma)\tag{2.6.41}$$

估计基准函数 $\theta(t)$, 其中

$$\bar{X}(t,\gamma)=\frac{\sum\limits_{i=1}^{n}\xi_i(t)\exp\{\gamma^{\mathrm{T}}X_i(t)\}X_i(t)}{\sum\limits_{i=1}^{n}\xi_i(t)\exp\{\gamma^{\mathrm{T}}X_i(t)\}},$$

$$\bar{Y}(t,\gamma)=\frac{\sum_{i=1}^{n}\xi_i(t)\exp\{\gamma^{\mathrm{T}}X_i(t)\}Y_i(t)}{\sum_{i=1}^{n}\xi_i(t)\exp\{\gamma^{\mathrm{T}}X_i(t)\}},$$

$\xi_i(t)=I(C_i\geqslant t)$. 在实际中, $\bar{X}(t,\gamma)$ 和 $\bar{Y}(t,\gamma)$ 可能是不可估的, 因为在删失时间之前不能完全观察到 $X(t)$ 和 $Y(t)$. 当观察不到他们的值时, Lin 和 Ying(2001) 用它们在近邻时间的相应值代替 $X(t)$ 和 $Y(t)$. 因此, 用 $\hat{\theta}(t;\beta,\gamma_0)$ 代替 (2.6.40) 式中的 $\theta(t)$ 可得到目标函数

$$Q_n(\beta,\gamma_0)=\sum_{i=1}^{n}\int_0^{\infty}w(t)[Y_i(t)-\beta^{\mathrm{T}}X_i(t)-\hat{\theta}(t;\beta,\gamma_0)]^2\mathrm{d}N_i(t). \tag{2.6.42}$$

当然, γ_0 是未知的, 但可以用它的矩估计 $\hat{\gamma}$ 来相合地估计它, 即 $\hat{\gamma}$ 是下列估计方程的解:

$$\sum_{i=1}^{n}\int_0^{\infty}[X_i(t)-\bar{X}_i(t;\gamma)]\mathrm{d}N_i(t)=0.$$

给定 $\hat{\gamma}$, 极小化 $Q_n(\beta,\hat{\gamma})$ 可以得到 β_0 的估计量, 记为 $\hat{\beta}$, 并称它为加权最小二乘估计. 估计量 $\hat{\beta}$ 等价于下列估计方程的解, 即 $\hat{\beta}$ 满足

$$\sum_{i=1}^{n}\int_0^{\infty}w(t)\{[X_i(t)-\bar{X}_i(t;\hat{\gamma})][Y_i(t)-\bar{Y}_i(t;\hat{\gamma})-\beta^{\mathrm{T}}[X_i(t)-\bar{X}_i(t;\hat{\gamma})]\}\mathrm{d}N_i(t)=0.$$

估计量 $\hat{\beta}$ 有明确的表达式

$$\begin{aligned}\hat{\beta}=&\left\{\sum_{i=1}^{n}\int_0^{\infty}w(t)[X_i(t)-\bar{X}_i(t;\hat{\gamma})]^{\otimes 2}\mathrm{d}N_i(t)\right\}^{-1}\\&\times\sum_{i=1}^{n}\int_0^{\infty}w(t)[X_i(t)-\bar{X}_i(t;\hat{\gamma})][Y_i(t)-\bar{Y}_i(t;\hat{\gamma})]\mathrm{d}N_i(t).\end{aligned}$$

将估计量 $\hat{\beta}$ 和 $\hat{\gamma}$ 代入到 (2.6.41) 式即可得到 $\theta(t)$ 的估计量 $\hat{\theta}^*(t)=\hat{\theta}(t;\hat{\beta},\hat{\gamma})$. Lin 和 Ying(2001) 证明了估计量 $\hat{\beta}$ 的渐近正态性, 并给出了 $\hat{\beta}$ 的标准误的估计量.

值得指出, Xue 和 Zhu(2007) 用经验似然方法研究了模型 (2.6.39). 他们构造了参数向量和基准函数的经验似然比统计量和极大经验似然估计量, 并证明了他们的渐近性质. 由此构造了兴趣参数置信域和置信区间. 限于篇幅, 这里不再赘述.

2.7　模 型 检 验

本节考虑模型 (2.1.1) 的检验问题, 其内容取材于 (Zhu and Ng, 2003). 不失一般性, 本节假定 X 的均值为 0. 当拟合独立观察数据 $\{(X_i,U_i,Y_i),1\leqslant i\leqslant n\}$ 时,

检验拟合模型 (2.1.1) 的适应性是重要的. 本节考虑的零假设为

$$H_0: E(Y|X=x,U=u)=\beta_0^{\mathrm{T}}x+g(u),\quad \text{对某个}\beta_0\text{和}g, \tag{2.7.1}$$

备择假设为

$$H_1: E(Y|X=x,U=u)\neq\beta_0^{\mathrm{T}}x+g(u),\quad \text{对某个}\beta_0\text{和}g.$$

记 $g_1(u)=E(Y|U=u)$, $g_2(u)=E(X|U=u)$. 在模型 (2.1.1) 两边求关于 U 的条件数学期望可得 $g(U)=g_1(U)-\beta_0^{\mathrm{T}}g_2(U)$. 由此可知 H_0 为真当且仅当

$$E(Y|X,U)=\beta_0^{\mathrm{T}}[X-g_2(U)]-g_1(U).$$

亦即, H_0 为真当且仅当

$$E\{[Y-\beta_0^{\mathrm{T}}(X-g_2(U))-g_1(U)]|X,U\}=0.$$

这就推出, 对所有 x, u,

$$E\{[Y-\beta_0^{\mathrm{T}}(X-g_2(U))-g_1(U)]w(U)I(X\leqslant x,U\leqslant u)\}=0, \tag{2.7.2}$$

其中 "$X\leqslant x$" 表示 X 的每一个分量小于或等于 x 的每一个分量, $w(u)$ 是任意的权函数. 通常取 $w(u)$ 为 (a,b) 上的示性函数, 其中 $0<a<b<1$. 因此, 在 H_0 下, (2.7.2) 式左边的经验版本

$$\frac{1}{n}\sum_{i=1}^{n}\{Y_i-\beta_0^{\mathrm{T}}[X_i-g_2(U_i)]-g_1(U_i)\}w(U_i)I(X_i\leqslant x,U_i\leqslant u)$$

将接近于 0. 为利用上式构造检验统计量, 需要估计 β_0, $g_1(\cdot)$ 和 $g_2(\cdot)$. 这里使用 2.2.1 小节给出的权函数法. 给定核函数 $K(\cdot)$ 和带宽 $h=h_n$, $g_1(U_i)$ 和 $g_2(U_i)$ 的核估计分别定义为

$$\hat{g}_1(U_i)=\sum_{j\neq i}^{n}W_{nj}(U_i)Y_j,\quad \hat{g}_2(U_i)=\sum_{j\neq i}^{n}W_{nj}(U_i)X_j,$$

其中

$$W_{nj}(U_i)=\frac{K_h(U_j-U_i)}{\sum_{j\neq i}^{n}K_h(U_j-U_i)},$$

$K_h(\cdot)=h^{-1}K(\cdot/h)$. β_0 的估计量定义为

$$\hat{\beta}=\hat{S}^{-1}\cdot\frac{1}{n}\sum_{i=1}^{n}[X_i-\hat{g}_2(U_i)][Y_i-\hat{g}_1(U_i)]w^2(U_i),$$

其中

$$\hat{S} = \frac{1}{n}\sum_{i=1}^{n}[X_i - \hat{g}_2(U_i)][X_i - \hat{g}_2(U_i)]^{\mathrm{T}} w^2(U_i).$$

考虑残差符号经验过程

$$R_n(X,U) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\varepsilon}_i w(U_i) I(X_i \leqslant x, U_i \leqslant u).$$

提出的检验统计量定义为

$$T_n = \int R_n^2(X,U)\mathrm{d}F_n(X,U), \tag{2.7.3}$$

其中 F_n 是基于样本 $\{(X_i,U_i), 1 \leqslant i \leqslant n\}$ 的经验分布. 我们将对 T_n 大的值拒绝零假设.

下面陈述统计量 R_n 和 T_n 的渐近性质. 记 $S_{g_2} = E\{[X - g_2(U)]^{\otimes 2} w(U)\}$,

$$\begin{aligned} J(X,U,Y;\beta_0,g_1,g_2,F,x,u) = &\big\{Y - \beta_0^{\mathrm{T}}[X - g_2(U)] - g_1(U)\big\}\big\{I(X \leqslant x, U \leqslant u) \\ &- E[I(X \leqslant x, U \leqslant u)(X - g_2(U))^{\mathrm{T}} w(U)] \\ &\times S_{g_2}^{-1}[X - g_2(U)] - F(X|U)I(U \leqslant u)\big\} w(U), \end{aligned}$$

其中 $F(x|u)$ 为给定 U 下 X 的条件分布函数. Zhu 和 Ng(2003) 在一些正则条件下证明了下述结论: 当 H_0 为真时, 有

$$R_n(x,u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} J(X_i,U_i,Y_i;\beta_0,g_1,g_2,F,x,u) + o_P(1), \tag{2.7.4}$$

且 $R_n(x,u)$ 在 Skorokhod 空间 $D[-\infty,\infty]^{p+1}$ 上依分布收敛于 $R(x,u)$, 其中 $R(x,u)$ 是中心连续 Gauss 过程, 其协方差函数为

$$\begin{aligned} &E[R(x_1,u_1)R(x_2,u_2)] \\ =\ &E[J(X,U,Y;\beta_0,g_1,g_2,F,x_1,u_1)J(X,U,Y;\beta_0,g_1,g_2,F,x_2,u_2)]. \end{aligned} \tag{2.7.5}$$

因此,

$$T_n \xrightarrow{D} \int R^2(X,U)\mathrm{d}F(X,U),$$

其中 $F(x,u)$ 为 (X,U) 的分布函数.

下面讨论检验对备择假设的敏感性. 考虑关于 n 的模型序列

$$E(Y|X,U) = \beta_0^{\mathrm{T}} X + g(U) + \psi(X,U)/\sqrt{n}, \tag{2.7.6}$$

其中 $\psi(X,U)$ 具有零均值, 且满足条件: 存在一个原点的邻域 $\mathcal{T}$ 和常数 $c>0$, 使得对任何 $t\in\mathcal{T}$, 有

$$|E[\psi(X,U)|U=u+t]-E[\psi(X,U)|U=u]|\leqslant c|t|,\quad \forall u.$$

Zhu 和 Ng(2003) 在一些正则条件下证明了下述结论: 在 (2.7.6) 的备择假设下, R_n 依分布收敛于 $R(x,u)+\psi^*(x,u)$, 其中

$$\begin{aligned}\psi^*(x,u)=&E\{[\psi(X,U)-E(\psi(X,U)|U)]w(U)I(X\leqslant x,U\leqslant u)\}\\&-E\{[X-g_2(U)]^{\mathrm{T}}[\psi(X,U)-E(\psi(X,U)|U)]w^2(U)\}\\&\times S^{-1}E\{[X-g_2(U)]w(U)I(X\leqslant x,U\leqslant u)\}\end{aligned}$$

是一个非随机位移函数. 因此,

$$T_n\xrightarrow{D}\int[R(X,U)+\psi^*(X,U)]^2\mathrm{d}F(X,U),$$

由 $\psi^*(x,u)$ 的表达式可以看到: 它不等于 0, 除非 $\psi^*(X,U)=\beta_0^{\mathrm{T}}X$. 因此, 检验统计量 T_n 可以从零假设检测任意接近 $n^{-1/2}$ 的局部备择假设.

应用上述结果可以进行模型检验, 具体步骤如下.

步骤 1. 产生均值为 0 方差为 1 的独立随机变量 $e_1,\cdots,e_n$. 记 $E_n=(e_1,\cdots,e_n)$, 定义 $R_n(x,u)$ 的条件对应为

$$R_n(E_n,x,u)=\frac{1}{\sqrt{n}}\sum_{i=1}^n e_iJ(X_i,U_i,Y_i;\hat\beta,\hat g_1,\hat g_2,\hat F,x,u),\tag{2.7.7}$$

其中 $\hat\beta$, $\hat g_1$, $\hat g_2$ 和 $\hat F$ 是 $R_n(x,u)$ 中的未知量的相合估计. 由此产生的条件检验统计量为

$$T_n(E_n)=\int R_n^2(E_n,x,u)\mathrm{d}F_n(x,u).\tag{2.7.8}$$

步骤 2. 产生 E_n 的 m 个集合, 记为 $E_n^{(i)},i=1,\cdots,m$, 从而得到 $T_n(E_n)$ 的 m 个值, 并记为 $T_n(E_n^{(i)}),i=1,\cdots,m$.

步骤 3. 估计的 p 值为 $\hat p=k/(m+1)$, 其中 k 是 $T_n(E_n^{(i)})$ 中大于或等于 T_n 的数目. 对给定的显著性水平 α, 当 $\hat p\leqslant\alpha$ 时拒绝 H_0.

Zhu 和 Ng(2003) 证明了如下结果: 在对 H_0 和 H_1 以及其他一些条件下, 对几乎所有的样本序列 $\{(X_1,U_1,Y_1),\cdots,(X_n,U_n,Y_n),\cdots\}$, $R_n(E_n,x,u)$ 的条件分布收敛到 $R_n(x,u)$ 的渐近零分布.

$T_n(E_n)$ 的条件分布用作确定检验的 p 值. 无论数据是来自假设的模型或是来自备择的模型, 我们自然希望这个条件分布能很好地逼近检验统计量的零分布. 换

句话说, 当应用再抽样方法时, 我们不知道数据的基本模型, 我们冒着在备择假设下条件分布背离检验的零分布的风险. 如果这样的话, p 值的确定可能是不正确的, 并且损失了功效的性能. 上述结果说明: 基于随机对称逼近的条件分布可能避免这个麻烦.

2.8 变量选择

变量选择是统计数据分析必不可缺少的工具. 从参数模型到半参数模型的变量选择, 已取得了许多成果. 本节主要介绍部分线性模型中变量选择的方法, 其内容主要取材于 (Fan and Li, 2004) 和 (Liang and Li, 2009) 等.

2.8.1 惩罚最小二乘

由模型 (2.1.1), 并注意到 $g(u) = g_1(u) - \beta_0^{\mathrm{T}} g_2(u)$, 其中 $g_1(u) = E(Y|U=u)$, $g_2(u) = E(X|U=u)$, 定义最小二乘函数

$$\frac{1}{2}\sum_{i=1}^{n}\left\{[Y_i - g_1(U_i)] - \beta^{\mathrm{T}}[X_i - g_2(U_i)]\right\}^2.$$

使用局部线性估计 $\hat{g}_1(\cdot)$ 和 $\hat{g}_2(\cdot)$ 分别代替 $g_1(\cdot)$ 和 $g_2(\cdot)$, 可得惩罚最小二乘函数

$$\mathcal{L}(\beta) = \frac{1}{2}\sum_{i=1}^{n}\left\{[Y_i - \hat{g}_1(U_i)] - \beta^{\mathrm{T}}[X_i - \hat{g}_2(U_i)]\right\}^2 + n\sum_{j=1}^{p} p_{\lambda_j}(|\beta_j|), \tag{2.8.1}$$

其中 $p_{\lambda_j}(\cdot)$ 可取任何一个惩罚函数. 比如 Fan 和 Li(2001) 提出的 SCAD 惩罚函数

$$p'_\lambda(\theta) = \lambda\left\{I(\theta \leqslant \lambda) + \frac{(a\lambda - \theta)_+}{(a-1)\lambda} I(\theta > \lambda)\right\}, \quad \theta > 0,$$

$p_\lambda(0) = 0$, $a = 3.7$. 讨厌参数 λ 可以用数据驱动方法选择. 极小化函数 $\mathcal{L}(\beta)$ 可得 β_0 的惩罚最小二乘估计量, 并记为 $\hat{\beta}$.

寻找 $\mathcal{L}(\beta)$ 的解是一个挑战性的问题, 因为诸如 $L_q(0 < q \leqslant 1)$ 惩罚或 SCAD 惩罚的惩罚函数 $p_{\lambda_j}(|\beta_j|)$ 在原点处是不正则的, 且在某点处可能没有二阶导数. 遵照 Fan 和 Li(2001) 的思想, 可以用二次函数局部地逼近惩罚函数, 具体做法如下. 给定接近 $\mathcal{L}(\beta)$ 的极小值的初始值 $\beta^{(0)}$ 和一个规定的值 η, 当 $|\beta^{(0)}| \geqslant \eta$ 时, 可以用二次函数局部地逼近惩罚函数 $p_{\lambda_j}(|\beta_j|)$, 即

$$[p_{\lambda_j}(|\beta_j|)]' = p'_{\lambda_j}(|\beta_j|)\mathrm{sgn}(\beta_j) \approx \left[p'_{\lambda_j}(|\beta_j^{(0)}|)/|\beta_j^{(0)}|\right]\beta_j.$$

利用二次逼近, Newton-Raphson 算法可以直接用来实施极小化 $\mathcal{L}(\beta)$. 此外, Newton-Raphson 算法其实是一个迭代的岭回归算法. 例如, 用

$$\sum_{i=1}^{n}[X_i - \hat{g}_2(U_i)][Y_i - \hat{g}_1(U_i)]$$

更新惩罚最小二乘的解, 其中

$$\Sigma_{\boldsymbol{\lambda}}(\beta^{(0)}) = \mathrm{diag}\{p'_{\lambda_1}(|\beta_1^{(0)}|)/|\beta_1^{(0)}|, \cdots, p'_{\lambda_p}(|\beta_p^{(0)}|)/|\beta_p^{(0)}|\},$$

$\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_p)$. 当有一个分量 $|\beta_j^{(0)}| < \eta$ 时, 将它置为 0. 在实施迭代中, 我们取不加惩罚的最小二乘估计作为初始值, 并反复更新 $\beta^{(0)}$.

当算法收敛时, 估计量 $\hat{\beta}$ 满足条件

$$\left.\frac{\partial l(\beta)}{\partial \beta_j}\right|_{\beta=\hat{\beta}} + np'_{\lambda_j}(|\hat{\beta}_j|)\mathrm{sgn}(\hat{\beta}_j) = 0, \tag{2.8.2}$$

称它为非零矩的惩罚最小二乘方程, 其中

$$l(\beta) = \frac{1}{2}\sum_{i=1}^{n}\left\{[Y_i - \hat{g}_1(U_i)] - \beta^{\mathrm{T}}[X_i - \hat{g}_2(U_i)]\right\}^2. \tag{2.8.3}$$

利用局部二次逼近, 迭代岭回归类似于 Newton-Raphson 算法. 因此, 可以用迭代岭回归获得系数估计的一个稳健的经验标准误 (SE) 公式. 换句话说, 对非零分量, 可以用 $n^{-1}[\hat{D} + \Sigma_{\boldsymbol{\lambda}}(\hat{\beta})]^{-1}\hat{V}[\hat{D} + \Sigma_{\boldsymbol{\lambda}}(\hat{\beta})]^{-1}$ 相合地估计 $\hat{\beta}$ 的协方差阵, 其中

$$\hat{D} = \frac{1}{n}\sum_{i=1}^{n}[X_i - \hat{g}_2(U_i)][X_i - \hat{g}_2(U_i)]^{\mathrm{T}}, \tag{2.8.4}$$

$$\hat{V} = \frac{1}{n}\sum_{i=1}^{n}[X_i - \hat{g}_1(U_i)][X_i - \hat{g}_2(U_i)]^{\mathrm{T}}\hat{\varepsilon}_i^2,$$

$\hat{\varepsilon}_i^2 = Y_i - \hat{g}_1(U_i) - \hat{\beta}^{\mathrm{T}}[X_i - \hat{g}_2(U_i)]$.

在实施前面描述的算法时, 我们希望有一个自动的数据驱动方法来估计讨厌参数 $\lambda_1, \cdots, \lambda_p$. 对非参数回归中的线性估计 (就响应变量而言), 有许多选择光滑参数的方法. 然而, 惩罚最小二乘估计不是线性的, 但借助于局部二次逼近, 它们是近似线性的. 因此, 可以使用诸如 CV 和 GCV 的各种光滑参数选择方法. 这里通过极小化近似的 GCV 得分来估计 $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_p)$. 回顾迭代岭回归算法可以用来得到惩罚最小二乘估计. 通过一些直接计算, 迭代岭回归算法最后步的有效参数数目为

$$e(\boldsymbol{\lambda}) = \mathrm{tr}\{[\widetilde{D} + \Sigma_{\boldsymbol{\lambda}}(\hat{\beta})]^{-1}\widetilde{D}\},$$

其中 $\widetilde{D}$ 是 $\hat{D}$ 的相应于 $\hat{\beta}$ 的非零分量的子矩阵, $\hat{D}$ 在 (2.8.4) 式中定义. 因此 GCV 统计量定义为

$$\mathrm{GCV}(\boldsymbol{\lambda}) = \frac{2l(\hat{\beta})}{n[1-e(\boldsymbol{\lambda})/n]^2},$$

其中 $l(\cdot)$ 在 (2.8.3) 式中定义. 我们选择 $\hat{\boldsymbol{\lambda}} = \arg\min\limits_{\boldsymbol{\lambda}}\{\mathrm{GCV}(\boldsymbol{\lambda})\}$.

为寻找最优的 $\boldsymbol{\lambda}$, 我们需要在 p 维空间上极小化 $\mathrm{GCV}(\boldsymbol{\lambda})$. 这是一个相当困难的任务. 直观地, 认为 λ_j 的大小将与 β_j 的惩罚最小二乘估计的标准误成比例. 因此, 在实际中可以令 $\boldsymbol{\lambda}_j = \lambda\,\mathrm{se}(\hat{\beta}_{\mathrm{LS}})$, $j = 1, \cdots, p$, 其中 $\mathrm{se}(\hat{\beta}_{\mathrm{LS}})$ 表示不加惩罚的 profile 最小二乘估计的 SE. 因此, 可以在一维空间上极小化 $\mathrm{GCV}(\lambda)$, 这将节省大量费用.

为给出估计量的渐近性质, 首先引入下列记号. 假设惩罚函数是非负且非减的, $p_{\lambda_j}(0) = 0$. 又设

$$a_n = \max_j\{p'_{\lambda_j}(|\beta_{j0}|)|\,\beta_{j0} \neq 0\},$$

$$b_n = \max_j\{p''_{\lambda_j}(|\beta_{j0}|)|\,\beta_{j0} \neq 0\}.$$

下面的定理给出了 $\hat{\beta}$ 的收敛速度, 其证明见文献 (Fan and Li, 2004).

定理 2.8.1 设 $h_n = cn^{-\alpha}$, $\dfrac{1}{8} < \alpha < \dfrac{1}{2}$, $c > 0$ 为某个常数, $a_n \to 0$, $b_n \to 0$, 则以概率 1 存在 $\mathcal{L}(\beta)$ 的局部最小值点 $\hat{\beta}$, 使得 $||\hat{\beta} - \beta_0|| = O_P(n^{-1/2} + a_n)$.

定理 2.8.1 说明了惩罚最小二乘估计 $\hat{\beta}$ 的收敛速度是如何依赖于 λ_j, 为了达到 $\sqrt{n}$ 收敛速度, 我们必须取足够小的 λ_j 以致 $a_n = O_P(n^{-1/2})$. 下面建立 $\hat{\beta}$ 的 oracle 性质. 假设 β_S 是由 β_0 的所有非零向量构成, β_N 是由 β_0 的所有零向量构成. 令

$$\beta_0^{\mathrm{T}}X = \beta_S^{\mathrm{T}}X_S + \beta_N^{\mathrm{T}}X_N = \beta_S^{\mathrm{T}}X_S,$$

其中 X_S 和 X_N 是协变量的两个子集. 上式右边第一部分为显著部分, 而第二部分为不显著部分. 为表达方便, 不失一般性, 我们假定 β_0 的前 s 个分量都不等于 0, 其他分量都等于 0. 即 $\beta_{10} = \beta_S$, $\beta_{20} = \beta_N$. 记

$$\Sigma = \mathrm{diag}\{p''_{\lambda_1}(|\beta_{10}|), \cdots, p''_{\lambda_s}(|\beta_{s0}|)\},$$

$$\boldsymbol{b} = (p'_{\lambda_1}(|\beta_{10}|)\mathrm{sgn}(|\beta_{10}|), \cdots, p'_{\lambda_s}(|\beta_{10}|)\mathrm{sgn}(|\beta_{s0}|))^{\mathrm{T}},$$

进一步, 假设 $\hat{\beta}_1$ 是由 $\hat{\beta}$ 的前 s 个分量组成, $\hat{\beta}_2$ 是由 $\hat{\beta}$ 的后 $p-s$ 个分量组成. $A = E\{[X - g_2(U)]^{\otimes 2}\}$, $B = E\{[X - g_2(U)]^{\otimes 2}\varepsilon^2\}$, A_{11} 和 B_{11} 由 A 和 B 的前 s 行和列组成.

Fan 和 Li(2004) 考虑了纵向数据部分线性模型, 在一些正则条件下, 证明了估计量的 oracle 性质. 这个性质对截面数据仍成立, 即有下列结果.

定理 2.8.2(oracle 性质) 设 $\lambda_j \to 0$, $\sqrt{n}\lambda_j \to \infty$, 惩罚函数 $p_{\lambda_j}(|\beta_j|)$ 满足

$$\liminf_{n\to\infty} \liminf_{\beta_j\to 0+} \frac{p'_{\lambda_j}(\beta_j)}{\lambda_j} > 0.$$

如果 $a_n = O_P(n^{-1/2})$, 则在定理 2.8.1 中的条件下, 定理 2.8.1 中 $\sqrt{n}$ 相合的估计量 $\hat{\beta} = (\hat{\beta}_1^{\mathrm{T}}, \hat{\beta}_2^{\mathrm{T}})^{\mathrm{T}}$ 必定满足下列 oracle 性质:

a. (稀疏性) $\hat{\beta}_2 = 0$;

b. (渐近正态性)

$$\sqrt{n}(A_{11} + \Sigma)[\hat{\beta}_1 - \beta_{10} + (A_{11} + \Sigma)^{-1}\boldsymbol{b}] \xrightarrow{D} N_s(0, B_{11}),$$

其中 $N_s(0, B_{11})$ 表示 s 维正态分布.

定理 2.8.2 蕴涵: 本节的方法正确地指定真模型和估计回归系数犹如我们知道子模型一样有效. 如果使用 SCAD 惩罚函数, 那么当 n 充分大时, $a_n = 0$, 因此得到的估计具有 oracle 性质. 然而, LASSO 惩罚不具有 oracle 性质, 因为不能同时满足条件 $a_n = \max_j \lambda_j = O_P(n^{-1/2})$ 和 $\sqrt{n}\lambda_j \to \infty$.

2.8.2 惩罚分位回归

在一些问题中, 最小二乘方法可能不理想, 此时分位回归可取而代之. 下面提出分位回归变量选择的方法. 定义 τ 分位目标函数

$$\rho_\tau(u) = \tau \max\{u, 0\} + (1-\tau)\max\{-u, 0\}.$$

注意到关于 $t \in \mathbf{R}$ 极小化 $E[\rho_\tau(\varepsilon - t)]$ 的解是 ε 的 τ 分位数. 因此, 惩罚分位回归函数具有形式

$$\mathcal{L}_T(\beta) = \sum_{i=1}^{n} \rho_\tau\big([Y_i - \hat{g}_1(U_i)] - \beta^{\mathrm{T}}[X_i - \hat{g}_2(U_i)]\big) + n\sum_{j=1}^{p} p_{\lambda_j}(|\beta_j|), \tag{2.8.5}$$

其中 $\hat{g}_1(\cdot)$ 和 $\hat{g}_2(\cdot)$ 分别为 $g_1(\cdot)$ 和 $g_2(\cdot)$ 的局部线性估计, $p_{\lambda_j}(\cdot)$ 为 SCAD 惩罚函数. 极小化 $\mathcal{L}_T(\beta)$ 即得到 β_0 的惩罚分位回归估计, 并记为 $\hat{\beta}_\tau$. 可以证明: 在适当的条件下, $\hat{\beta}_\tau$ 是 $\sqrt{n}$ 相合的并具有 oracle 性质. 详见文献 (Liang and Li, 2009).

2.8.3 惩罚多项式样条

本节使用多项式样条方法构造 β_0 的惩罚估计, 其基本思想如下. 我们用多项式样条逼近 $g(\cdot)$. 对给定的正整数 M_n, 将 $[0,1]$ 分成 $M_n + 1$ 个子区间 $I_{n\nu} = [\xi_{n\nu}, \xi_{n,\nu+1})$, $\nu = 0, 1, \cdots, M_n - 1$, $I_{nM_n} = [\xi_{nM_n}, 1]$, 其中 $\xi_{n0} = 0$, $\xi_{n,M_n+1} = 1$. 用 $\Delta_n = \{\xi_{n\nu}\}_{\nu=1}^{M_n}$ 表示该分割. 假定 $\max_{0\leqslant\nu\leqslant M_n}\{\xi_{n,\nu+1} - \xi_{n\nu}\} = O(M_n)$. 设 $\mathcal{S}_m(\Delta_n)$ 为

在点 $\xi_{n1}, \cdots, \xi_{nM_n}$ 处有简单节的 m 阶多项式样条空间. 这个空间包含具有下列两个性质的所有函数 $s(\cdot)$:

(i) $s(\cdot)$ 是限制在任何区间 $I_{n\nu}(0 \leqslant \nu \leqslant M_n)$ 上的 m 阶多项式;

(ii) 若 $m \geqslant 2$, 那么 $s(\cdot)$ 在 $[0,1]$ 上是 $m-2$ 次连续可微的.

按照文献 (Schumaker, 1981) 的推论 2.10, 存在 $\mathcal{S}_m(\Delta_n)$ 的一个局部基 $\{B_{nl}(u), 1 \leqslant l \leqslant q_n\}$, 其中 $q_n = M_n + m$ 是 $\mathcal{S}_m(\Delta_n)$ 的维数. 记 $Z_n(u) = (B_{n1}(u), \cdots, B_{nq_n}(u))^{\mathrm{T}}$. 可以将 $\mathcal{S}_m(\Delta_n)$ 中的任何元素 $s(u)$ 写为 $s(u) = \gamma^{\mathrm{T}} Z_n(u)$, 其中 γ 为 $q_n \times 1$ 向量. 我们试图在 $\mathcal{S}_m(\Delta_n)$ 中寻找接近于 $g(\cdot)$ 的 $s(\cdot)$. 在适当的光滑条件下, $\mathcal{S}_m(\Delta_n)$ 中的元素可以很好地逼近 $g(\cdot)$. 因此, 估计 $g(\cdot)$ 的问题就转化为估计 γ 的问题.

利用 SCAD 惩罚, 定义惩罚最小二乘目标函数

$$\mathcal{L}_{\mathrm{PS}}(\beta, \gamma) = \frac{1}{2} \sum_{i=1}^{n} \left\{ [Y_i - \beta^{\mathrm{T}} X_i - \gamma^{\mathrm{T}} Z_n(U_i)] \right\}^2 + n \sum_{j=1}^{p} p_{\lambda_j}(|\beta_j|). \tag{2.8.6}$$

令 $(\hat{\beta}, \hat{\gamma}) = \arg\min\limits_{\beta,\gamma} \mathcal{L}_{\mathrm{PS}}(\beta, \gamma)$. 于是, β_0 和 $g(u)$ 的惩罚多项式样条估计可分别定义为 $\hat{\beta}_{\mathrm{PS}} = \hat{\beta}$ 和 $\hat{g}_{\mathrm{PS}}(u) = \hat{\gamma}^{\mathrm{T}} Z_n(u)$.

Xie 和 Huang(2009) 研究了高维部分线性模型的 SCAD 惩罚问题. 得到了参数和非参数分量的惩罚估计的收敛速度, 证明了非零系数的 SCAD 惩罚估计量具有渐近 oracle 性质.

参考文献

柴根象, 洪圣岩. 1995. 半参数回归模型. 合肥: 安徽教育出版社

柴根象, 施云驰, 钱志坚. 2001. 固定设计的时间序列半参数回归. 数学年刊, 22A(2): 163~176

柴根象, 孙平. 1995. 半参数回归模型的二阶段估计的渐近正态性. 应用数学学报, 18(3): 354~363

柴根象, 徐克军. 1999. 半参数回归的线性小波光滑. 应用概率统计, 15(1): 97~105

崔恒建. 1995. 半参数 EV 模型的参数估计理论. 科学通报, 40(16): 1444~1447

高集体, 陈希孺, 赵林城. 1994. 部分线性模型中估计的渐近正态性. 数学学报, 37(2): 256~268

高集体, 洪圣岩, 梁华. 1995. 部分线性模型中估计的收敛速度. 数学学报, 38(5): 658~669

洪圣岩. 1991. 一类半参数回归模型的估计理论. 中国科学, A 辑, 21(12): 1258~1271

洪圣岩, 成平. 1993. 半参数回归模型的参数估计的 Bootstrap 逼近. 中国科学, A 辑, 23(3): 239~251

洪圣岩, 赵忠柏. 1993. 偏线性模型的核 ~- 最小二乘估计法的渐近性质. 数学年刊, 14A(6): 717~731

胡舒合. 1994. 固定设计下半参数回归模型估计的强相合性. 数学学报, 37(3): 393~401

刘强, 薛留根. 2009. 纵向数据下部分线性 EV 模型的渐近性质. 应用数学学报, 32(1): 178~189

刘强, 薛留根, 陈放. 2009. 删失数据下部分线性 EV 模型中参数的经验似然置信域. 数学学报, 52(3): 549~560

钱伟民, 柴根象. 1999. 半参数回归模型小波估计的强逼近. 中国科学, 29(3): 233~240

钱伟民, 柴根象, 蒋凤英. 2000. 半参数回归模型的误差方差的小波估计. 数学年刊, 21A(3): 341~350

石坚. 1998. 部分线性模型中的 Edgeworth 展开. 数学学报, 41(4): 683~686

孙孝前, 尤进红. 2003. 纵向数据半参数建模中的迭代加权偏样条最小二乘估计. 中国科学, 33(5): 470~480

王启华. 1995. 随机截断下半参数回归模型的相合估计. 中国科学, 25(8): 819~832

王启华, 郑忠国. 1997. 随机删失下半参数回归模型中估计的渐近性质. 中国科学, 27(7): 583~594

薛留根. 1999. 随机删失下半参数回归模型的估计理论. 数学年刊, 20(6): 745~754

薛留根. 2002. 半参数回归模型随机加权 M 估计的强逼近. 应用数学学报, 25(4): 591~603

薛留根. 2003. 半参数回归模型中小波估计的随机加权逼近速度. 应用数学学报, 26(1): 11~25

薛留根. 2004. 半参数回归模型中误差方差估计之分布的非一致收敛速度. 应用数学学报, 27(2): 379~381

薛留根. 2005. 半参数回归模型中误差方差估计的 Berry-Esseen 界. 数学学报, 48(1): 157~170

薛留根, 田萍. 2004. 固定设计下半参数回归模型的随机加权逼近. 北京工业大学学报, 30(2): 247~251

薛留根, 朱力行. 2007. 纵向数据下部分线性模型的经验似然推断. 中国科学, A 辑, 37(1): 31~44

尤进红, 陈歌迈. 2004. 带有异方差的部分线性回归模型的 B 样条估计. 数学年刊, 25(5): 661~676

Ai C R, You J H, Zhou Y. 2011. Statistical inference using a weighted difference based series approach for partially linear regression models. Journal of Multivariate Analysis, 102(3): 601~618

Aneiros-Pérez G, Vilar-Fernández J M. 2008. Local polynomial estimation in partial linear regression models under dependence. Computational Statistics & Data Analysis, 52(5): 2757~2777

Arnold S F. 1981. The Theory of Linear Models and Multivariate Analysis. New York: John Wiley & Sons

Avramidis P. 2005. Two-step cross-validation selection method for partially linear models. Statistica sinica, 15(4): 1033~1048

Bai Y, Zhu Z Y, Fung W K. 2008. Partial linear models for longitudinal data based on quadratic inference functions. Scandinavian Journal of Statistics, 35(1): 104~118

Bhattacharya P K, Zhao P L. 1997. Semiparametric inference in a partial linear model. The Annals of Statistics, 25(1): 244~262

Bianco A, Boente G, González-Manteiga W, et al. 2010. Estimation of the marginal location under a partially linear model with missing responses. Computational Statistics & Data Analysis, 54(2): 546~564

Bianco A, Boente G, Martinez E. 2006. Robust tests in semiparametric partly linear models. Scandinavian Journal of Statistics, 33(3): 435~450

Boente G, He X M, Zhou J H. 2006. Robust estimates in generalized partially linear models. The Annals of Statistics, 34(6): 2856~2878

de Boor C. 2001. A practical guide to splines, Revised Edition. New York: Springer

Bunea F. 2004. Consistent covariate selection and post model selection inference in semiparametric regression. The Annals of Statistics, 32: 898~927

Bunea F, Wegkamp M. 2004. Two-stage model selection procedures in partially linear regression. The Canadian Journal of Statistics, 32: 105~118

Chen H. 1988. Convergence rates for parametric components in a partly linear model. The Annals of Statistics, 16: 136~146

Chen H, Shiau J H. 1991. A two-stage spline smoothing method for partially linear models. Journal of Statistical Planning and Inference, 27: 187-201

Chen H, Shiau J H. 1994. Data-driven efficient estimators for a partially linear model. The Annals of Statistics, 22(1): 211~237

Chen M, Yuen K C, Zhu L X. 2003. Asymptotics of the goodness-of-fit test for a partial linear model with randomly censored data. Science in China, Series A, 46: 145~158

Chen S N, Khan S. 2001. Semiparametric estimation of a partially linear censored regression model. Econometric Theory, 17(3): 567~590

Cuzick J. 1992. Semiparametric additive regression. Journal of the Royal Statistical Society, Series B, 54(3): 831~843

Dabo-Niang S, Guillas S. 2010. Functional semiparametric partially linear model with autoregressive errors. Journal of Multivariate Analysis, 101(2): 307~315

Dette H, Marchlewski M. 2008. A test for the parametric form of the variance function in a partial linear regression model. Journal of Statistical Planning and Inference, 138: 3005~3021

Ding H J, Claeskens G, Jansen M. 2011. Variable selection in partially linear wavelet models. Statistical Modelling, 11(5): 409~427

Dong C H, Gao J, Tong H. 2007. Semiparametric penalty function method in partially linear model selection. Statistica Sinica, 17(1): 99~114

Engle R F, Granger C W J, Rice J, et al. 1986. Semiparametric estimates of the relation

between weather and electricity scales. Journal of the American Statistical Association, 81: 310~320

Eubank R L, Kambour E L, Kim J T, et al. 1998. Estimation in partially linear models. Computational Statistics & Data Analysis, 29: 27~34

Fan J, Gijbels I. 1996. Local Polynomial Modeling and its Applications. London: Chapman and Hall

Fan J, Li R Z. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96: 1348~1360

Fan J, Li R Z. 2004. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. Journal of the American Statistical Association, 99: 710~723

Fu Y Z, Chen X D. 2012. Model selection of generalized partially linear models with missing covariates. Journal of Statistical Planning and Inference, 142(1): 126~138

Gai Y J, Lin L, Wang X L. 2011. Consistent inference for biased sub-model of high-dimensional partially linear model. Journal of Statistical Planning and Inference, 141(5): 1888~1898

Gao J T. 1995. Asymptotic theory to partly linear models. Communications in Statistics. Theory and Methods, 24: 1985~2010

Gao J T, Shi P D. 1997. M-type smoothing splines in nonparametric and semiparametric regression models. Statistica Sinica, 7(4): 1155~1169

Gao J T, Tong H. 2004. Semiparametric non-linear time series model selection. Journal of the Royal Statistical Society, Series B, 66: 321~336

Gao J T, Zhao L C. 1993. Adaptive estimation in partly linear regression models. Science in China, Series A, 23(1): 14~27

Gasser T, Müller H G. 1979. Kernel estimation of regression function. In Smoothing Techniques for Curve Estimation, Lecture Notes in Mathematics, 757: 23~68. New York: Springer-Verlag

Green P. 1985. Linear model for field trials, smoothing and cross-validation. Biometrika, 527~537

Green P, Silverman B W. 1994. Nonparametric regression and generalized linear models: a Roughness Penalty Approach. New York: Chapman and Hall

Hall P, Titterington D M. 1988. On confidence bands in nonparametric density estimation and regression. Journal of Multivariate Analysis, 27: 228~254

Hamilton S A, Truong Y K. 1997. Local linear estimation in partly linear models. Journal of Multivariate Analysis, 60: 1~19

Härdle W, Liang H, Gao J T. 2000. Partially Linear Models. Heidelberg: Physica Verlag

Härdle W, Mammen E, Müller M. 1998. Testing parametric versus semiparametric modeling in generalized linear models. Journal of the American Statistical Association, 93:

1461~1474

He X M, Fung W K, Zhu Z Y. 2005. Robust estimation in generalized partial linear models for clustered data. Journal of the American Statistical Association, 100(472): 1176~1184

He X M, Liang H. 2000. Quantile regression estimates for a class of linear and partially linear errors-in-variables models. Statistica Sinica, 10(1): 129~140

He X M, Zhu Z Y, Fung W K. 2002. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika, 89(3): 579~590

Heckman N. 1986. Spline smoothing in a partly linear model. Journal of the Royal Statistical Society, Series B, 48: 244~248

Hong S Y, Cheng P. 1999. Normal approximation rate of the kernel smoothing estimator in a partial linear model. Journal of Multivariate Analysis, 70(2): 207~220

Hu Z H, Wang N Y, Carroll R J. 2004. Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data. Biometrika, 91(2): 251~262

Huang J Z, Zhang L Y, Zhou L. 2007. Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. Scandinavian Journal of Statistics, 34(3): 451~477

Huang Z S. 2012. Empirical likelihood for the parametric part in partially linear errors-in-function models. Statistics & Probability Letters, 82(1): 63~66

Huber P J. 1981. Robust Statistics. New York: Wiley

Kendall M, Stuart A. 1992. The Advanced Theory of Statistics 2, 4th ed. London: Griffin

Leng C L, Liang H, Martinson N. 2011. Parametric variable selection in generalized partially linear models with an application to assess condom use by HIV-infected patients. Statistics in Medicine, 30(16): 2015~2027

Li N, Xu X Z, Jin P. 2011. Testing the linearity in partially linear models. Journal of Nonparametric Statistics, 23(1): 99~114

Liang H. 1992. Asymptotic Efficiency in Semiparametric Models and Related Topics. Ph D Thesis. Institute of Systems Science, Chinese Academy of Sciences, Beijing, P R China

Liang H. 2000. Asymptotic normality of parametric part in partially linear models with measurement error in the nonparametric part. Journal of Statistical Planning and Inference, 86(1): 51~62

Liang H. 2006. Estimation in partially linear models and numerical comparisons. Computational Statistics & Data Analysis, 50(3): 675~687

Liang H. 2008. Generalized partially linear models with missing covariates. Journal of Multivariate Analysis, 99(5): 880~895

Liang H. 2009. Generalized partially linear mixed-effects models incorporating mismeasured covariates. Annals of the Institute of Statistical Mathematics, 61(1): 27~46

Liang H, Härdle W, Carroll R J. 1999. Estimation in a semiparametric partially linear errors-in-variables model. The Annals of Statistics, 27(5): 1519~1535

Liang H, Härdle W, Sommerfeld V. 2000. Bootstrap approximation in a partially linear regression model. Journal of Statistical Planning and Inference, 91(2): 413~426

Liang H, Li R Z. 2009. Variable selection for partially linear models with measurement errors. Journal of the American Statistical Association, 104(485): 234~248

Liang H, Qin Y S. 2008. Empirical likelihood-based inferences for partially linear models with missing covariates. Australian & New Zealand Journal of Statistics, 50(4): 347~359

Liang H, Qin Y S, Zhang X Y, et al. 2009. Empirical likelihood-based inferences for generalized partially linear models. Scandinavian Journal of Statistics, 36(3): 433~443

Liang H, Thurston S W, Ruppert D, et al. 2008. Additive partial linear models with measurement errors. Biometrika, 95(3): 667~678

Liang H, Wang S J, Carroll R J. 2004. Partially linear models with missing response variables and error-prone covariates. Biometrika, 94: 185~198

Liang H, Wang S J, Robins J M, et al. 2004. Estimation in partially linear models with missing covariates. Journal of the American Statistical Association, 96(466): 357~367

Liang H, Wang S J, Robins J M, et al. 2007. Estimation in partially linear models with missing covariates. Journal of the American Statistical Association, 99(466): 357~367

Liang H Y, Jing B Y. 2009. Asymptotic normality in partial linear models based on dependent errors. Journal of Statistical Planning and Inference, 139(4): 1357~1371

Lin D Y, Ying Z. 2001. Semiparametric and nonparametric regression analysis of longitudinal data (with Discussion). Journal of the American Statistical Association, 96: 103~126

Lin X H, Carroll R J. 2001a. Semiparametric regression for clustered data using generalized estimating equations. Journal of the American Statistical Association, 96: 1045~1056

Lin X H, Carroll R J. 2001b. Semiparametric regression for clustered data. Biometrika, 88(4): 1179~1185

Little R J A, Rubin D B. 1987. Statistical Analysis with Missing Data. New York: Wiley

Liu L. 2007. Estimation of generalized partially linear models with measurement error using sufficiency scores. Statistics & Probability Letters, 77(15): 1580~1588

Liu Q. 2011. Asymptotic normality for the partially linear EV models with longitudinal data. Communications in Statistics—Theory and Methods, 40(7): 1149~1158

Liu X H, Wang Z Z, Hu X M. 2011. Testing heteroscedasticity in partially linear models with missing covariates. Journal of Nonparametric Statistics, 23(2): 321~337

Lu X W. 2009. Empirical likelihood for heteroscedastic partially linear models. Journal of Multivariate Analysis, 100(3): 387~396

Ma Y Y, Carroll R J. 2006. Locally efficient estimators for semiparametric models with mea-

surement error. Journal of the American Statistical Association, 101(476): 1465~1474

Ma Y Y, Chiou J M, Wang N. 2006. Efficient semiparametric estimator for heteroscedastic partially linear models. Biometrika, 93(1): 75~84

Mammen E, van de Geer S. 1997. Penalized quasi-likelihood estimation in partial linear models. The Annals of Statistics, 25: 1014~1035

Mao J, Zhu Z Y. 2011. Joint mean-covariance models with applications to longitudinal data in partially linear model. Communications in Statistics-Theory and Methods, 40(17): 3119~3140

Muller S, Vial C. 2009. Partially linear model selection by the bootstrap. Australian & New Zealand Journal of Statistics, 51(2): 183~200

Ni H F. 2012. Penalized least-squares estimation for regression coefficients in high-dimensional partially linear models. Journal of Statistical Planning and Inference, 14(2): 379~389

Ni X, Zhang H H, Zhang D W. 2009. Automatic model selection for partially linear models. Journal of Multivariate Analysis, 100(9): 2100~2111

Opsomer J D, Ruppert D. 1997. Fitting a bivariate additive model by local polynomial regression. The Annals of Statistics, 25: 186~211

Owen A B. 1990. Empirical likelihood ratio confidence regions. The Annals of Statistics, 18(1): 90~120

Owen A B. 1991. Empirical Likelihood for Linear Models. The Annals of Statistics, 19(4): 1725~1747

Pitrun I, King M L, Zhang X B. 2006. Smoothing spline based tests for non-linearity in a partially linear model. Journal of Statistical Planning and Inference, 136(8): 2446~2469

Qin G Y, Zhu Z Y. 2008. Robust estimation in partial linear mixed model for longitudinal data. Acta Mathematica Scientia, 28(2): 333~347

Qin G Y, Zhu Z Y, Fung W K. 2009. Robust estimation of covariance parameters in partial linear model for longitudinal data. Journal of Statistical Planning and Inference, 139(2): 558~570

Qin J, Lawless J. 1994. Empirical likelihood and general estimating equations. The Annals of Statistics, 22: 300~325

Qin Y S, Li J J. 2011a. Empirical likelihood for partially linear models with missing responses at random. Journal of Nonparametric Statistics, 23(2): 497~511

Qin Y S, Li Y H. 2011b. Empirical likelihood for partially linear models with missing responses: the fixed design case. Communications in Statistics—Theory and Methods, 40(10): 1849~1865

Rao J N K, Scott A J. 1981. The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. Journal of the American Statistical Association, 76: 221~230

Rice J A. 1984. Boundary modification for kernel regression. Communications in Statistics—Theory and Methods, 13: 893～900

Rice J A. 1986. Covergence rates for partial linear models. Statistics & Probability Letters, 4: 203～208

Rice J A, Silverman B W. 1991. Estimating the mean and covariance structure nonparametrically when the data are cures. Journal of the Royal Statistical Society, Series B, 53: 233～243

Robinson P M. 1988. Root-n-consistent semiparametric regression. Econometrika, 56: 931～954

Robins J M, Rotnitzky A, Zhao L P. 1994. Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association, 89: 846～866

Schimek M G. 2000. Estimation and inference in partially linear models with smoothing splines. Journal of Statistical Planning and Inference, 91(2): 525～540

Schumaker L. 1981. Spline Functions: Basic Theory. New York: Wiley

Serfling R. 1980. Approximation Theorems of Mathematical Statistics. New York: Wiley

Severini T A, Staniswalis J G. 1994. Quasilikelihood estimation in semiparametric models. Journal of the American Statistical Association, 89: 501～511

Shao Q. 2009. Seasonality analysis of time series in partial linear models. Journal of Nonparametric Statistics, 21(7): 827～837

Shen C W, Tsou T S, Balakrishnan N. 2011. Robust likelihood inference for regression parameters in partially linear models. Computational Statistics & Data Analysis, 55(4): 1696～1714

Shi J, Lau T S. 2000. Empirical likelihood for partially linear models. Journal of Multivariate Analysis, 41: 425～433

Shi P D. 1992. M Estimation for Partly Linear Models. Ph D Thesis, Institute of Systems Science, Chinese Academy of Sciences, Beijing, P R China

Speckman P. 1988. Kernel smoothing in partial linear models. Journal of the Royal Statistical Society, Series B, 50: 413～436

Spiegelman C, Sacks J. 1980. Consistent window estimation in nonparametric regression. The Annals of Statistics, 5: 595～620

Stone C J. 1980. Optimal rates of convergence for nonparametric estimators. The Annals of Statistics, 8: 1348～1360

Sun Z H, Wang Q H, Dai P J. 2009. Model checking for partially linear models with missing responses at random. Journal of Multivariate Analysis, 100: 636～651

Wang N. 2003. Marginal nonparametric kernel regression accounting for within subject correlation. Biometrika, 90: 43～52

Wang Q H. 1996. Consistent estimators in random censorship semiparametric regression

models. Science in China, Series A, 39: 163~176

Wang Q H. 2009. Statistical estimation in partial linear models with covariate data missing at random. Annals of the Institute of Statistical Mathematics, 61(1): 47~84

Wang Q H, Li G. 2002. Empirical likelihood semiparametric regression analysis under random censorship. Journal of Multivariate Analysis, 83(2): 469~486

Wang Q H, Linton O, Härdle W. 2004. Semiparametric regression analysis with missing response at random. Journal of the American Statistical Association, 99: 334~345

Wang Q H, Jing B Y. 1999. Empirical likelihood for partial linear models with fixed designs. Statistics & Probability Letters, 41(4): 425~433

Wang Q H, Jing B Y. 2003. Empirical likelihood for partial linear models. Annals of The Institute of Statistical Mathematics, 55(3): 585~595

Wang Q H, Sun Z H. 2007. Estimation in partially linear models with missing responses at random. Journal of Multivariate Analysis, 98(7): 1470~1493

Wang Q H, Yu K M. 2007. Likelihood-based kernel estimation in semiparametric errors-in-covariables models with validation data. Journal of Multivariate Analysis, 98(3): 455~480

Wang Q H, Zheng Z G. 1997. Asymptotic properties for the semiparametric regression model with randomly censored data. Science in China, Series A, 40: 945~957

Xie H L, Huang J. 2009. SCAD-penalized regression in high-dimensional partially linear models. The Annals of Statistics, 37: 673~696

Xue H Q, Lam K F, Li G Y. 2004. Sieve maximum likelihood estimator for semiparametric regression models with current status data. Journal of the American Statistical Association, 99(466): 346~356

Xue L G, Liu Q. 2010. Bootstrap approximation of wavelet estimates in a semiparametric regression model. Acta Mathematica Sinica-English Series, 26(4): 763~778

Xue L G, Xue D. 2011. Empirical likelihood for semiparametric regression model with missing response data. Journal of Multivariate Analysis, 102(4): 723~740

Xue L G, Zhu L X. 2005. L_1-norm estimation and random weighting method in a semiparametric model. Acta Mathematicae Applicatae Sinica, 21(2): 295~302

Xue L G, Zhu L X. 2007. Empirical likelihood semiparametric regression analysis for longitudinal data. Biometrika, 1~16

Xue L G, Zhu L X. 2008. Empirical likelihood-based inference in a partially linear model for longitudinal data. Science in China, Series A, (English Version), 51(1): 115~130

Yang Y P, Xue L G, Cheng W H. 2009. Empirical likelihood for a partially linear model with covariate data missing at random. Journal of Statistical Planning and Inference, 139(12): 4143~4153

Yatchew A. 1997. An Elementary estimator of the partial linear model. Economics Letters, 57: 135~143

You J H, Chen G M. 2005. Testing heteroscedasticity in partially linear regression models. Statistics & Probability Letters, 73: 61~70

You J H, Xu Q F, Zhou B. 2008. Statistical inference for partially linear regression models with measurement errors. Chinese Annals of Mathematics, Series B, 29(2): 207~222

You J H, Zhou X, Zhou Y. 2011. Series Estimation in Partially Linear In-Slide Regression Models. Scandinavian Journal of Statistics, 38(1): 89~107

You J H, Zhou X. 2005. Bootstrap of a semiparametric partially linear model with autoregressive errors. Statistica Sinica, 15(1): 117~133

You J H, Zhou X. 2009. Partially linear models and polynomial spline approximations for the analysis of unbalanced panel data. Journal of Statistical Planning and Inference, 139(3): 679~695

Zeger S L, Diggle P J. 1994. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics, 50: 689~699

Zhang D, Lin X H, Raz J, et al. 1998. Semiparametric stochastic mixed models for longitudinal data. Journal of the American Statistical Association, 93(442): 710~719

Zhang J H, Feng S Y, Li G R. 2011. Empirical likelihood inference for partially linear panel data models with fixed effects. Economics Letters, 113(2): 165~167

Zhao H B, You J H. 2011. Difference based estimation for partially linear regression models with measurement errors. Journal of Multivariate Analysis, 102(10): 1321~1338

Zhou H H, Zhu Z Y, Fung W K. 2008. Robust testing with generalized partial linear models for longitudinal data, Journal of Statistical Planning and Inference, 138(6): 1871~1883

Zhu L X, Ng K W. 2003. Checking the adequacy of a partial linear model. Statistica Sinica, 13: 763~781

Zhu Z Y, He X M, Fung W K. 2003. Local influence analysis for penalized Gaussian likelihood estimators in partially linear models. Scandinavian Journal of Statistics, 30(4): 767~780

第 3 章 单指标模型

3.1 引 言

单指标模型首先考虑 p 维协变量 X 的线性组合, 把所有的协变量投影到一个线性空间上, 然后在这个一维线性空间上拟合一个一元函数. 假设 $\{(X_i, Y_i), 1 \leqslant i \leqslant n\}$ 是来自 (X, Y) 的 iid 样本, 那么单指标模型具有形式

$$Y_i = g(\beta_0^{\mathrm{T}} X_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{3.1.1}$$

其中 $g(\cdot)$ 为一元未知联系函数, β_0 为 p 维指标系数, ε_i 为 iid 随机误差, 且几乎处处 $E(\varepsilon_i|X_i) = 0$. 为了模型的可识别性, 需要假定 $||\beta_0|| = 1$, 且 β_0 的第一个非 0 元素是正的, 其中 $||\cdot||$ 表示 Euclid 模. Ichimura(1993) 和 Horowitz(1998) 研究了单指标模型的可识别性, Manski(1988) 讨论了二元响应模型的可识别性. 由于指标 $\beta_0^{\mathrm{T}} X$ 合并了 X 的维数, 把 p 维协变量降维到一元指标, 从而使得单指标模型避免了多元非参数回归中出现的“维数灾祸”问题.

单指标模型是 20 世纪 80 年代中后期发展起来的一种重要的统计模型. 该模型在金融经济和生物医学等领域具有广泛的应用背景. 研究单指标模型的一大任务就是对联系函数和指标系数进行估计. 关于 β_0 的估计方法, 已有不少学者进行了研究. 概括而言, 根据是否需要求解非线性最优化问题, 可以把 β_0 的估计方法分为两大类: M 估计法和直接估计法. 其中 β_0 的 M 估计法包括半参数最小二乘估计法和半参数极大似然估计法, 直接估计法包括平均导数法和切片逆回归法. 此外, 近年来又提出了几种新的估计方法, 其中包括: 最小平均导数估计法、惩罚样条估计法、估计方程估计法和经验似然估计法等.

Ichimura(1993) 和 Härdle 等 (1993) 提出了半参数最小二乘估计方法. 该方法采用了参数回归模型中最小二乘估计的思想. 在指标系数给定的情况下, 用去掉一点的 N-W 估计方法来估计联系函数 $g(\cdot)$, 所得的估计是指标系数 β_0 的一个函数, 然后用非线性最小二乘的思想极小化残差平方和, 可得指标系数 β_0 的估计. 由于该方法需要解决复杂的非线性最优化问题, 因此计算较为复杂. Weisberg 和 Welsh(1994) 用类似的方法研究了模型 (3.1.1), 提出了指标系数的估计量, 并在一般的条件下证明了它的 $\sqrt{n}$ 相合性.

半参数极大似然估计沿用了参数极大似然估计的思想. 在单指标模型中, X 和 Y 的联合分布以及给定 X 下 Y 的条件密度都既依赖于参数 β_0 又依赖于联系函数

$g(\cdot)$. Delecroix 等 (2003) 证明了 β_0 的半参数极大似然估计是渐近有效的, 保持了参数极大似然估计的最重要的性质. Delecroix 等 (2006) 对单指标模型分析了一大类半参数 M 估计, 其中包括半参数拟似然和半参数极大似然估计, 并证明了所提出的估计量的渐近正态性.

虽然 M 估计具有有效性, 渐近正态性和自动带宽选取等许多优点, 但该方法的最大缺点是它需要求解复杂的最优化问题. 尽管直接估计方法的理论性质不如 M 估计, 但由于其能够提供估计量的解析形式, 因此, 它在单指标模型的估计中具有一定的吸引力. 平均导数方法由 Stoker(1986) 与 Härdle 和 Stoker(1989) 提出. 随后, Powell 等 (1989) 提出了密度加权平均导数估计方法, 该方法不需要迭代并且容易计算, 还适用于 X 的所有分量都是连续型随机变量的情形. Härdle 和 Tsybakov(1993) 证明了平均导数方法导致一个 $\sqrt{n}$ 相合的估计量. Horowitz 和 Härdle(1996) 和 Hristache 等 (2001) 把这种方法进一步推广到 X 的分量有离散随机变量的情形. 其他相关的工作有 (Newey and Stoker, 1993; Chaudhuri et al., 1997; Xia, 2007, 2008) 等.

Li(1989,1991) 提出了切片逆回归方法来削减输入变量的维数, 从而得到了降维方向的估计. 切片逆回归估计量也能达到 $\sqrt{n}$ 相合. Yin 和 Cook(2005) 结合似然、相关、逆回归和信息理论的思想提出了一个一般的降维方法. 一些相关的工作包括 (Duan and Li, 1991; Hsing and Carroll, 1992; Zhu and Ng, 1995; Zhu and Fang,1996; Naik and Tsai, 2000; Prendergast, 2005) 等.

Xia 等 (2002b) 提出了最小平均方差估计方法来构造多指标模型中降维空间的估计. 该方法在没有对联系函数进行欠光滑的情况下也能使参数的估计达到更快的收敛速度. Xia(2006) 证明了平均导数估计和最小平均方差估计的渐近正态性. Xia,Tong 和 Li(2002a) 考虑了一个单指标可变模型, 提出了指标系数和联系函数的估计量, 并证明了它们的渐近性质.

尽管半参数方法有好的理论性质, 但对指标系数进行估计时它需要解高维极小化问题, 这就带来计算上的困难. 现有方法的数值计算上的不稳定性愈加注重带宽选择或迭代改进技巧; 见文献 (Yu and Ruppert, 2002; Carroll et al., 1997). Bayes 方法提供了更稳定估计的希望, 尤其对具有低信噪比的小或中等数据集. 所付出的代价是需要计算精细的 Markov chain Monte-Carlo(MCMC) 技术. 尽管如此, 单指标模型的简单形式蕴涵了 Bayes 方法将相对容易发展和执行. Antoniadis 等 (2004) 发展了包括频率方法在内的 Bayes 方法: 联系函数用 B 样条逼近, 指标系数的先验信息取作 Fisher-von Mises, 且采用广义交叉验证准则避免联系函数的过拟合. 一个随机游动 Metropolis 算法被用作抽取后验样本. Wang(2009) 用 Bayes 方法研究了单指标模型的参数估计和变量选择问题. 他设计了一个新的随机游动 Metropolis 采样器来抽取指标系数的条件后验分布的样本.

Wu 等 (2010) 研究了单指标分位数回归模型的估计和推断问题. 他们提出了针对这个模型的算法, 那就是基于局部线性光滑来估计非参数部分, 线性分位数回归来解指标系数部分. 利用这个算法能相当方便地估计单指标模型中的参数分量和非参数分量, 得到了非参数联系函数的估计, 条件中位数的估计和指标系数的估计. 他们也证明了这些估计量的渐近性质. 条件分位数的置信区间也是欣然可用的. 此外, 他们也得到了一个便于计算的拇指规则带宽选择器. 最近, Kong 和 Xia(2011) 研究了广义单指标模型的分位数估计问题.

Chiou 和 Müller(1998) 利用拟似然方法研究了单指标模型, 其中模型误差是异方差的, 且方差函数是未知的. 他们提出了三阶段方法来估计联系函数、方差函数和指标系数, 得到了指标系数估计量的渐近分布以及联系函数和方差函数估计量的收敛速度. 通过比较在联系函数和方差函数已知情形下得到的参数拟似然估计, 证明了提出的估计量是渐近有效的. 他们也提出了基于 Pearson χ^2 统计量的数据适应的带宽选择方法, 通过模拟研究证明了这种带宽选择方法产生好的效果. Cui 等 (2011) 也将拟似然方法应用到了异方差单指标模型的研究, 并发展了这个方法. 他们提出的估计函数方法 (EFM) 是通过解估计方程来得到指标系数的估计, 并证明了它们的相合性和渐近正态性. 与现有方法相比较, 他们提出的估计量有小的渐近方差. 他们也提出了一个迭代算法来计算指标系数的估计.

Naik 和 Tsai(2000) 研究了单指标模型中偏最小二乘估计的性能, 尤其预测变量呈现高的共线性. Kulasekera 和 Lin(2010) 提出了单指标模型的误差方差的两个估计量, 并就估计的优良性与现有估计量进行了比较. Jiang 和 Wang(2011) 提出了纵向和函数响应数据的单指标模型, 利用指标系数的恰当的初始估计, 证明了所提出的估计量是 $\sqrt{n}$ 相合的并具有渐近正态分布. 他们也考虑了未知联系函数的非参数估计并给出了该估计量的最优收敛速度. 该方法的优点是可以用同一带宽来估计非参数均值函数和指标系数. 关于单指标模型中兴趣参数的估计问题的其他研究成果可参见 (Huh and Park, 2002; Zhou and Liang, 2006; Lin and Kulasekera, 2007, Zhu and Zhu, 2009; Choi et al., 2011; Leitenstorfer and Tutz, 2011; Lin, 2011; Wang et al., 2011) 等.

Xue 和 Zhu(2006) 研究了模型 (3.1.1) 的经验似然推断问题. 一个估计的经验似然比统计量用来构造指标系数的置信域. 他们也提出了一个调整的经验似然比统计量, 并证明了它渐近于 χ^2 变量. Zhu 和 Xue(2006) 用经验似然方法研究了部分线性单指标模型. 他们提出了一种 "纠偏" 方法使构造的经验似然比服从 χ^2 分布. 纠偏的经验似然方法比最小二乘法有一些明显的特征: 一是不需要方差估计; 二是估计模型中非参数函数时避免了欠光滑. 纠偏方法对纯粹的单指标模型仍使用. 作为副产品, 纠偏方法也可以应用到其他半参数回归模型中. 见文献 (Xue and Zhu, 2007a, 2007b; Xue, 2009a, 2009b, 2010; Huang and Zhang, 2011; Xue and Xue, 2011;

Xue et al., 2011) 等.

Xia 等 (2004) 利用 bootstrap 法研究了单指标模型的拟合优度检验问题. 他们提出了一个 "纠偏" 方法来处理检验统计量中的偏差问题. 因为他们要使用 bootstrap 方法模拟检验统计量的零分布, 所以他们进一步使用它通过对 bootstrap 值的简单平均来估计偏差项. 这样做不需要欠光滑或过光滑联系函数. 因此, 他们的方法容易执行且完全由数据驱动. Stute 和 Zhu(2005) 研究了某些得分类检验统计量的大样本性质. 此外, 他们提出了基于拟合优度检验的特征函数. 单指标模型检验的其他研究成果见文献 (Härdle et al., 1997; Xia, 2009; Escanciano and Song, 2010; Klein and Shen, 2010; Zhang et al., 2010) 等.

Kong 和 Xia(2007) 研究了模型 (3.1.1) 的变量选择问题. 他们指出在线性回归模型或非参数模型中常用的去掉 m 个个体的交叉验证法在单指标模型中不能得到很好的使用. 为了克服这一难点, 他们提出了一种新方法, 即分离交叉验证法 (separated cross-validation method), 有效地解决了单指标模型中的变量选择问题. 关于单指标模型的变量选择的其他研究成果见文献 (Naik and Tsai, 2004; Wang, 2009; Zhu et al., 2011; Peng and Huang, 2011) 等.

最近, 利用单指标模型对复杂数据进行统计分析也取得了一些成果. Bai 等 (2009) 利用惩罚样条和二次推断函数方法研究了纵向数据的单指标模型, 证明了提出的估计量的渐近性质. Lin 和 Kulasekera(2010) 在纵向数据下研究了线性单指标模型的同等问题. Lopez(2009) 研究了右删失响应下的单指标模型, 对 M 估计方法进行了推广, 并证明了提出的估计量的相合性和渐近正态性. Lu(2010) 对随机删失样本的单指标模型提出了两个估计方法, 构造了指标系数的估计量并证明了它们的 $\sqrt{n}$ 相合性和渐近正态性. Wang 等 (2010b) 在缺失响应下研究了单指标模型中兴趣参数的估计问题, 证明了提出的估计量的大样本性质.

3.2 半参数最小二乘估计

对每一个常数 c, 条件方差

$$\mathrm{var}(Y|\beta^{\mathrm{T}}X=c)=E\{[Y-E(Y|\beta^{\mathrm{T}}X)]^2|\beta^{\mathrm{T}}X=c\}$$

可以被用来度量 Y 在等高线 $\beta^{\mathrm{T}}X=c$ 上的变化趋势, 此处和下文假定 β 是 p 维单位向量. 因此, 估计 β_0 的一个合理的方法是首先构造 $E\{[Y-E(Y|\beta^{\mathrm{T}}X)]^2\}$ 的一个样本形式作为目标函数, 然后极小化这个目标函数而得到 β_0 的一个估计. 由于 $E(Y|\beta^{\mathrm{T}}X)$ 是未知的, 需要利用它的一个非参数回归估计代替它. 这里使用 N-W 核估计. 尽管 N-W 核估计在各种方法中不是最佳的, 但由于它计算简便且较容易

实施, 因而在单指标回归模型的研究中常常使用. 对该方法的详细论述可参见文献 (Härdle et al., 1993).

设 $A \subseteq \mathbf{R}^p$ 是一个选择的集合, 使得核估计公式中的分母不能太接近于 0. 定义

$$g(u|\beta) = E(Y|\beta^{\mathrm{T}} X_A = u),$$

其中 X_A 在 $X \in A$ 上与 X 有相同的分布. 为了消除估计的偏差, 用分离一点 N-W 估计 $\hat{g}^{(-i)}(\beta^{\mathrm{T}} X_i|\beta)$ 作为 $g(\beta^{\mathrm{T}} X_i|\beta)$ 的估计, 即

$$\hat{g}^{(-i)}(\beta^{\mathrm{T}} X_i|\beta) = \frac{\sum\limits_{j \neq i} Y_j K((\beta^{\mathrm{T}} X_j - \beta^{\mathrm{T}} X_i)/h)}{\sum\limits_{j \neq i} K((\beta^{\mathrm{T}} X_j - \beta^{\mathrm{T}} X_i)/h)}.$$

因此, β_0 的半参数最小二乘估计 $\hat{\beta}$ 定义为目标函数

$$Q_n(\beta) = \frac{1}{n} {\sum_i}' [Y_i - \hat{g}^{(-i)}(\beta^{\mathrm{T}} X_i|\beta)]^2$$

的最小值点, 其中 ${\sum\limits_i}'$ 表示对指标 i 求和使得 $X_i \in A$.

对同方差的情况, 即 $\mathrm{var}(\varepsilon_i|X_i) = \sigma^2$, Härdle 等 (1993) 证明了 $\hat{\beta}$ 具有渐近正态性, 即 $\sqrt{n}(\hat{\beta} - \beta_0)$ 的渐近分布为 $N(0, \sigma^2 W_1^-)$, 其中

$$\begin{aligned} W_1 = &\int_A [x - E(X_A|\beta_0^{\mathrm{T}} X_A = \beta_0^{\mathrm{T}} x)][x - E(X_A|\beta_0^{\mathrm{T}} X_A = \beta_0^{\mathrm{T}} x)]^{\mathrm{T}} \\ &\times g'(\beta_0 x)^2 f(x) \mathrm{d}x, \end{aligned} \tag{3.2.1}$$

$f(x)$ 为 X 的密度函数, W_1^- 表示 W_1 的广义逆.

即使当误差 ε_i 是异方差时, 估计量 $\hat{\beta}$ 也是 β_0 的相合估计 (Härdle et al., 1993). 然而, 在异方差情况下, 估计量 $\hat{\beta}$ 的有效性可以在定义 $Q_n(\beta)$ 时引入一个适当的权函数 $w(\cdot)$ 来改进. Ichimura(1993) 使用一个确定的光滑参数研究了这个问题. 下面我们概述包含 $w(\cdot)$ 时的最优光滑, 并且讨论 $w(\cdot)$ 的经验估计.

下文假定误差方差 $\sigma^2(x)$ 实际上是 $\beta_0^{\mathrm{T}} X$ 的函数, 在这种情况下可适当的取 $w(\cdot)$ 也是 $\beta_0^{\mathrm{T}} X$ 的函数. 使用权函数有它的优缺点. 除了增加计算的复杂性外 (尤其用经验确定), 权函数修改了 $Q_n(\beta)$ 的定义, 在某种程度上是不必要的. 然而, 这个问题可以使用多种方法解决. 因此, β_0 的半参数加权最小二乘估计 $\hat{\beta}^*$ 定义为目标函数

$$Q_n^*(\beta) = \frac{1}{n} {\sum_i}' [Y_i - \hat{g}^{(-i)}(\beta^{\mathrm{T}} X_i|\beta)]^2 w(X_i)$$

的最小值点. 正如 Härdle 等 (1993) 指出: 只要 $w(\cdot)$ 是有界且连续的正函数, 估计量 $\hat{\beta}^*$ 仍具有渐近正态性; 即当 $w(x)=\sigma^{-2}(x)$ 且 $\sigma^2(x)$ 仅是 $\beta_0^{\mathrm{T}}X$ 的函数时, $\sqrt{n}(\hat{\beta}^*-\beta_0)$ 的渐近分布为 $N(0,W_2^-)$, 其中

$$W_2=\int_A[x-E(X_A|\beta_0^{\mathrm{T}}X_A=\beta_0^{\mathrm{T}}x)][x-E(X_A|\beta_0^{\mathrm{T}}X_A=\beta_0^{\mathrm{T}}x)]^{\mathrm{T}} \\ \times g'(\beta_0x)^2w^2(x)\sigma^2(x)f(x)\mathrm{d}x.$$

在实际中, 方差函数 $\sigma^2(x)$ 通常是未知的, 需要对它进行估计. 我们主要考虑

$$\sigma^2(x)=\tau^2G(g(\beta_0^{\mathrm{T}}x)),$$

其中 $G(\cdot)$ 是一个已知的光滑函数, τ 是一个常数 (可能未知). 两阶段方法步骤如下:

(I) 极小化目标函数 $Q_n(\beta)$ 得到 β_0 的一个估计 $\hat{\beta}$;

(II) 取 $w(X_i)=G(\hat{g}^{(-i)}(\hat{\beta}^{\mathrm{T}}X_i|\hat{\beta}))^{-1}$, 极小化目标函数 $Q_n^*(\beta)$ 得到 β_0 的估计 $\hat{\beta}^*$.

可以证明: 如果$G(\cdot)$是一个二次可微函数,$\inf\limits_u G(u)>0$, 且取$w(x)=G(g(\beta_0^{\mathrm{T}}x))^{-1}$, 则 $\sqrt{n}(\hat{\beta}^*-\beta_0)$ 依分布收敛于正态分布 $N(0,W_3^-)$, 其中 W_3 是将 (3.2.1) 中的 $f(x)$ 改为 $f(x)\tau^{-2}G(g(\beta_0^{\mathrm{T}}x))^{-1}$ 而得到.

一般地, 参数 β_0 的 M 估计具有如下形式:

$$\hat{\beta}=\arg\min_{\beta\in R^p}\frac{1}{n}\sum_{i=1}^n\Psi(Y_i,\hat{g}^{(-i)}(\beta^{\mathrm{T}}X_i|\beta)),$$

其中 Ψ 为满足一定正则条件的 $\mathbf{R}^2\to\mathbf{R}$ 上的函数.

3.3 半参数极大似然估计

半参数极大似然估计沿用了参数极大似然估计的思想. 假设在给定 X 下 Y 的条件密度 $f(y|x)$ 存在, 并通过指标 $\beta_0^{\mathrm{T}}X$ 依赖于 X. 我们也假定 (X,Y) 的联合密度依赖于 β_0, 而 X 的边缘密度 $f(x)$ 不依赖于 β_0. 因此有一个函数 f_{β_0}, 满足

$$f(y|x)=f_{\beta_0}(y|\beta_0^{\mathrm{T}}x),\quad (x,y)\in\mathcal{Z}, \tag{3.3.1}$$

其中 $\mathcal{Z}$ 为 (X,Y) 的支撑集. 因此, 似然函数可表示为

$$L_n^*(\beta)=\prod_{i=1}^n f_\beta(Y_i|\beta^{\mathrm{T}}X_i)f(X_i).$$

对数似然函数为

$$l_n^*(\beta) = \sum_{i=1}^{n} \log f_\beta(Y_i|\beta^{\mathrm{T}} X_i) + \sum_{i=1}^{n} \log f(X_i).$$

易见, 上式最后一项不依赖于 β 和 $g(\cdot)$, 极大化 $l^*(\beta)$ 等于极大化上式右端的第一项. 从而可以定义 β 的对数似然函数

$$l_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \log f_\beta(Y_i|\beta^{\mathrm{T}} X_i) I\{(X_i, Y_i) \in S\},$$

其中 S 是 $\mathcal{Z}$ 的某个子集. 由于 $f_\beta(\cdot|\cdot)$ 是未知的, 需要用其估计量代替它. 为此, 定义 $f_\beta(Y_i|\beta^{\mathrm{T}} X_i)$ 的估计量

$$\hat{f}_\beta(Y_i|\beta^{\mathrm{T}} X_i) = \frac{N_{i,n}(Y_i, \beta^{\mathrm{T}} X_i)}{D_{i,n}(\beta^{\mathrm{T}} X_i)},$$

其中

$$N_{i,n}(Y_i, \beta^{\mathrm{T}} X_i) = \frac{1}{n-1} \sum_{j \neq i} K\left(\frac{Y_j - Y_i}{h}\right) K\left(\frac{\beta^{\mathrm{T}} X_j - \beta^{\mathrm{T}} X_i}{h}\right),$$

$$D_{i,n}(\beta^{\mathrm{T}} X_i) = \frac{1}{n-1} \sum_{j \neq i} K\left(\frac{\beta^{\mathrm{T}} X_j - \beta^{\mathrm{T}} X_i}{h}\right),$$

$h = h_n$ 为带宽, $K(\cdot)$ 为一个固定的核函数. 因此, 当 $f_\beta(\cdot|\cdot)$ 是未知时, β 的对数似然函数定义为

$$\hat{l}_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_\beta(Y_i|\beta^{\mathrm{T}} X_i) I\{(X_i, Y_i) \in S\}.$$

则可定义 β_0 的半参数极大似然估计为下列极大化问题的解:

$$\hat{\beta} = \arg\max_\beta \hat{l}_n(\beta). \tag{3.3.2}$$

Delecroix 等 (2003) 证明了下面两个结果:

$$\sup_{\beta \in \mathcal{B}} |\hat{l}_n(\beta) - l_n(\beta)| \to 0, \quad \text{a.s.},$$

$$\sup_{\beta \in \mathcal{B}} |l_n(\beta) - l(\beta)| \to 0, \quad \text{a.s.},$$

其中 $\mathcal{B} \subset \mathbf{R}^p$, $l(\beta) = E[\log f_\beta(Y|\beta^{\mathrm{T}} X) I\{(X, Y) \in S\}]$. 从上述结果可以得到: 如果 $l(\beta)$ 在 β_0 处有唯一的极大值, 那么 $\hat{\beta}$ 几乎处处收敛于这个极大值.

Delecroix 等 (2003) 也证明了 β 的半参数极大似然估计 $\hat{\beta}$ 的渐近正态性和渐近有效性, 它保持了参数极大似然估计的最重要的性质.

3.4 平均导数估计

Stoker(1986) 与 Härdle 和 Stoker(1989) 提出的平均导数方法与 M 估计法是平行的, 其基本思想是估计梯度向量

$$\nabla m(x) \equiv \frac{\partial m(x)}{\partial x} = \beta_0 g'(\beta_0^{\mathrm{T}} x). \tag{3.4.1}$$

的期望值, 其中 $m(x)$ 是 $\mathbf{R}^p$ 上的未知函数, 且满足

$$m(x) = g(\beta_0^{\mathrm{T}} X). \tag{3.4.2}$$

由此可以得到 β_0 的估计量 $\hat{\beta}$. 利用 $\hat{\beta}$ 及一元非参数回归的估计方法可以得到联系函数 $g(\cdot)$ 的估计量.

Härdle 和 Tsybakov(1993) 证明了 $\hat{\beta}$ 的相合性, 并讨论了 $g(\cdot)$ 的估计的大样本性质. 但是, 这个方法需要先获得 X 的边际密度的核估计, 然后获得 $E[\nabla m(x)]$ 的估计, 这在实际应用中是比较复杂的. Horowitz 和 Härdle(1996) 将平均导数估计推广到 X 的分量是连续和 (或) 离散的情形. 然而, 对平均导数估计的要求条件在应用上有其局限性. 尤其, 仅对于具有非常光滑的设计密度的随机设计才能获得速度 $n^{-1/2}$, 并且需要高阶核. 在实际应用中, 由于数据稀少而使用高阶核的设计密度估计, 效果会很差, 这就是所谓的 "维数灾祸" 问题.

Hristache 等 (2001) 引入了 β_0 的一个新的方向估计. 它可以看作平均导数估计量的迭代改良. 这个方法的一个重要特征是完全适应于未知光滑性的联系函数. 下面叙述这个方法. 从 (3.4.1) 式可以看出, 回归函数 $m(x)$ 的梯度 $\nabla m(x)$ 在每一点 x 处与 β_0 成比例, 这就诱发估计平均梯度的思想. 记

$$\gamma_0 \equiv \frac{1}{n}\sum_{i=1}^{n} \nabla m(X_i) = \frac{1}{n}\sum_{i=1}^{n} g'(\beta_0^{\mathrm{T}} X_i)\beta_0. \tag{3.4.3}$$

显然, γ_0 是未知回归函数的线性函数, 因此人们可以应用线性函数估计的成熟理论 (Ibragimov and Khasminski, 1987). 在实施这个方法时发生的主要问题是梯度向量不光滑, 并且关于设计矩阵和联系函数的苛刻假定必须成立以确保相应估计量的 $\sqrt{n}$ 相合性 (Samarov, 1993; Hädle and Tsybakov, 1993). 向量 γ_0 和 β_0 自然可以分别用下面两式估计:

$$\hat{\gamma}_1 = \frac{1}{n}\sum_{i=1}^{n} \widehat{\nabla m}_1(X_i), \quad \hat{\beta}_1 = \frac{\hat{\gamma}_1}{||\hat{\gamma}_1||}, \tag{3.4.4}$$

其中 $\widehat{\nabla m}_1(X_i)$ 是 $\nabla m(X_i)$ 的引导估计. 估计 $\nabla m(X_i)$ 和 $m(X_i)$ 的一个常规方法

是局部最小二乘法 (Fan and Gijbels, 1996), 即

$$\begin{pmatrix} \hat{m}_1(X_i) \\ \widehat{\nabla m}_1(X_i) \end{pmatrix} = \arg \inf_{a\in\mathbf{R},\gamma\in\mathbf{R}^p} \sum_{j=1}^{n} \left(Y_j - a - \gamma^{\mathrm{T}} X_{ij}\right)^2 K\left(\frac{||X_{ij}||^2}{h^2}\right),$$

其中 $X_{ij} = X_j - X_i$, $h = h_n$ 为带宽, 核 $K(\cdot)$ 是正的函数且具有支撑 $[-1,1]$, 使得在以 X_i 为中心 h 为半径的邻域 $U_h(X_i)$ 外面的所有点 X_j 的权被取消. 这个二次最优化问题的解可表示为

$$\begin{pmatrix} \hat{m}_1(X_i) \\ \widehat{\nabla m}_1(X_i) \end{pmatrix} = \left[\sum_{j=1}^{n} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^{\mathrm{T}} K\left(\frac{||X_{ij}||^2}{h^2}\right)\right]^{-1} \times \sum_{j=1}^{n} Y_j \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} K\left(\frac{||X_{ij}||^2}{h^2}\right). \tag{3.4.5}$$

Hristache 等 (2001) 证明了下述结果: 依概率, 有

$$||\hat{\gamma}_1 - \gamma_0|| \leqslant c_1 h + \frac{||\xi||}{h\sqrt{n}}, \tag{3.4.6}$$

其中 ξ 是 $\mathbf{R}^p$ 上的正态随机变量, $E(\xi) = 0$, $E(||\xi||^2) \leqslant c_2$, 且 c_1 和 c_2 是确定的常数. (3.4.6) 式的右端包含两项, 第一项为确定的误差 (偏差), 它是由 $m(\cdot)$ 的局部逼近引起的误差. 这个误差与 h 成比例. 第二项是独立于 $m(\cdot)$ 的随机误差; 该项的主阶是 $(h\sqrt{n})^{-1}$. 选择 h 的阶为 $n^{-1/4}$ 使两项达到平衡, 由此可得

$$||\hat{\gamma}_1 - \gamma_0|| = O_P(n^{-1/4}).$$

类似地, 可以得到

$$||\hat{\beta}_1 - \beta_0|| = O_P(n^{-1/4}).$$

如果维数 $p > 4$, 这种情况会很糟糕. 这是因为数据的稀疏性. 为了提供 $p+1$ 个设计点以便在半径 h 的球内来计算局部线性逼近 (3.4.5), 人们将取阶为 $n^{-1/p}$ 的 h. 这就在 (3.4.6) 中引起偏差 $O_P(n^{-1/p})$. 因此, 对 $p > 4$, 这样一个估计量的精度为 $n^{-1/p}$. 当然, 这个收敛速度 $n^{-1/(4\vee p)}$ 比能够达到的速度 $n^{-1/2}$ 相差甚远. 幸运地是, 估计量 $\hat{\beta}_1$ 可以大大改进. 其思想是利用结构假定 (3.4.2) 来改进梯度估计.

下面是 Hristache 等 (2001) 提出的观点. 为了找到向量 β_0, 我们不需要完全知道向量 γ_0. 仅 γ_0 的方向是重要的. 因此, 如果 $\hat{\gamma}$ 是 γ_0 的估计量, 它能确保 $\hat{\gamma}$ 和它在 β_0 上的投影 $(\hat{\gamma}^{\mathrm{T}}\beta_0)\beta_0$ 的差别来的小. 此外, 如果 γ_0 不接近于 0, 那么 $\hat{\gamma}/||\hat{\gamma}||$ 可以合理地估计单位向量 β_0. 我们的目的是在一定程度上修正简单估计量 (3.4.5), 其方法是改进与 β_0 正交的方向上的估计量.

此时假定 β_0 已知, 需要估计 $\nabla m(x)=\beta_0 g'(\beta_0^{\mathrm{T}}x)$. 设回归函数 $m(x)$ 及梯度函数在正交于 β_0 的子空间内没有变化. 这就蕴涵在带 $S=\{x:||(x-X_i)^{\mathrm{T}}\beta_0||\leqslant\rho\}$ 内函数 $m(\cdot)$ 可以用一个线性函数很好地逼近, 其中 ρ 是很小的正数. 现在设 $\widehat{\nabla m}(X_i)$ 是 $\nabla m(X_i)$ 在带 $S(X_i)$ 上基于 $m(\cdot)$ 的局部线性逼近的估计量, 即

$$\begin{pmatrix}\hat{m}(X_i)\\ \widehat{\nabla m}(X_i)\end{pmatrix}=\left\{\sum_{j=1}^{n}\begin{pmatrix}1\\ X_{ij}\end{pmatrix}\begin{pmatrix}1\\ X_{ij}\end{pmatrix}^{\mathrm{T}}K\left(\frac{||\beta_0^{\mathrm{T}}X_{ij}||^2}{\rho^2}\right)\right\}^{-1}$$
$$\times\sum_{j=1}^{n}Y_j\begin{pmatrix}1\\ X_{ij}\end{pmatrix}K\left(\frac{||\beta_0^{\mathrm{T}}X_{ij}||^2}{\rho^2}\right).\tag{3.4.7}$$

通过平均得到相应的估计量 $\hat{\gamma}$ 和 $\tilde{\beta}$, 即

$$\hat{\gamma}=\frac{1}{n}\sum_{i=1}^{n}\widehat{\nabla m}(X_i),\quad \tilde{\beta}=\frac{\hat{\gamma}}{||\hat{\gamma}||}.$$

可以选择 (3.4.7) 式中的 ρ 充分小, 如取 $n^{-1/3}$, 不依赖于维数 p, 并且在 (几乎) 每一个带 $S(X_i)$ 上有足够的设计点. 可以证明在正交于 β_0 的方向上所定义的梯度估计量 $\widehat{\nabla m}(X_i)$ 的偏差的阶是 ρ^2. 此外, 适当选择 ρ 且在一些正则条件下, 估计量 $\tilde{\beta}$ 是渐近正态的且具有收敛速度 $n^{-1/2}$. 不幸的是, 我们无法实施这个估计量, 因为它包含目标向量 β_0. 一个自然的想法是用引导估计 $\hat{\beta}_1$ 代替 β_0. 这就得到了下面的迭代程序. 我们开始于来自 (3.4.4) 和 (3.4.5) 式的通常估计量 $\hat{\beta}_1$ 以及某个 $h=h_1$. 虽然这个估计非常粗糙, 但它提供了 β_0 的有用信息. 在下一步, 我们使用椭球窗 $\{x:S_2||x-X_i||\leqslant h_2\}$ 代替球形窗 $\{x:||x-X_i||\leqslant h_1\}$ 来更新梯度估计量 $\widehat{\nabla m}_2(X_i)$, 其中 $S_2=(I+\rho_2^{-2}\hat{\gamma}_1\hat{\gamma}_1^{\mathrm{T}})^{1/2}$, $\rho_2<\rho_1=1$, $h_2>h_1$. 换句话说, 我们在方向 $\hat{\gamma}_1$ 上收缩原来的窗 (因 $\rho_2<1$) 并在所有正交方向上延伸它们 (因 $h_2>h_1$). 由此得到

$$\begin{pmatrix}\hat{m}_2(X_i)\\ \widehat{\nabla m}_2(X_i)\end{pmatrix}=\left\{\sum_{j=1}^{n}\begin{pmatrix}1\\ X_{ij}\end{pmatrix}\begin{pmatrix}1\\ X_{ij}\end{pmatrix}^{\mathrm{T}}K\left(\frac{||S_2X_{ij}||^2}{h_2^2}\right)\right\}^{-1}$$
$$\times\sum_{j=1}^{n}Y_j\begin{pmatrix}1\\ X_{ij}\end{pmatrix}K\left(\frac{||S_2X_{ij}||^2}{h_2^2}\right),\tag{3.4.8}$$

这就得到了 γ_0 的一个估计量:

$$\hat{\gamma}_2=\frac{1}{n}\sum_{i=1}^{n}\widehat{\nabla m}_2(X_i).$$

我们每次在当前估计量 $\hat{\gamma}_k$ 的方向上压缩平均窗并在正交于 $\hat{\gamma}_k$ 的超平面上伸展它们; 继续这种方式直到最后的窗非常像正交于 β_0 的平坦的层次. 经过 $k=k(n)$ 次

重复后, 这个算法给出一个估计量 $\hat{\beta}=\hat{\gamma}_k/||\hat{\gamma}_k||$. 下面我们给出计算估计量的算法步骤.

1. 迭代程序

这个程序需要输入参数 $\rho_{\min}<\rho_1$ 和 $h_1<h_{\max}$, 以便在迭代过程中 ρ 通过因子 a_ρ 从 ρ_1 到 $\rho_{\min}$ 几何递减, 而 h 通过因子 a_h 从 h_1 到 $h_{\max}$ 几何递增. 这些参数的选择将在下面讨论. 这个算法如下:

(1) 设定初值: 指定参数 ρ_1, $\rho_{\min}$, a_ρ, h_1, $h_{\max}$, a_h, $k=1$, $\hat{\beta}_0=0$, $C_w=0.5$. 定义 $\bar{w}$ 为矩阵 $\bar{V}$ 的最小特征值的平方根, 即 $\bar{w}=\{\lambda_{\min}(\bar{V})\}^{1/2}$, 其中

$$\bar{V}=\frac{1}{E[K(\zeta^{\mathrm{T}}\zeta)]}E\left[\begin{pmatrix}1\\ \zeta\end{pmatrix}\begin{pmatrix}1\\ \zeta\end{pmatrix}^{\mathrm{T}}K(\zeta^{\mathrm{T}}\zeta)\right],$$

ζ 是球 $B_1=\{x\in\mathbf{R}^p:||x||\leqslant 1\}$ 上服从均匀分布的随机变量.

(2) 计算 $S_k=(I+\rho_k^{-2}\hat{\gamma}_{k-1}\hat{\gamma}_{k-1}^{\mathrm{T}})^{1/2}$.

(3) 对每个 $i=1,\cdots,n$, 计算矩阵 $\hat{V}_k(X_i)$, 这里

$$\hat{V}_k(X_i)=\frac{1}{\sum\limits_{j=1}^{n}K(||W_{ij,k}||^2)}\sum_{j=1}^{n}\begin{pmatrix}1\\ W_{ij,k}\end{pmatrix}\begin{pmatrix}1\\ W_{ij,k}\end{pmatrix}^{\mathrm{T}}K(||W_{ij,k}||^2),$$

且 $W_{ij,k}=h_k^{-1}S_k(X_j-X_i)$. 定义 w_i 为矩阵 $\hat{V}_k(X_i)$ 的最小特征值的平方根, 即 $w_i=\{\lambda_{\min}(\hat{V}_k(X_i))\}^{1/2}$.

(4) 如果条件

$$w_1+\cdots+w_n\geqslant nC_w\bar{w}$$

不成立, 那么通过因子 a_h 增加 h_k, 即 $h_k:=a_hh_k$. 如果 $h_k>h_{\max}$, 结束, 否则继续步骤 (3).

(5) 对每一个 $i=1,\cdots,n$, 利用下式计算 $\widehat{\nabla m}_k(X_i)$:

$$\begin{pmatrix}\hat{m}_k(X_i)\\ \widehat{\nabla m}_k(X_i)\end{pmatrix}=\left\{\sum_{j=1}^{n}\begin{pmatrix}1\\ X_{ij}\end{pmatrix}\begin{pmatrix}1\\ X_{ij}\end{pmatrix}^{\mathrm{T}}K\left(\frac{||S_kX_{ij}||^2}{h_k^2}\right)\right\}^{-1}\times\sum_{j=1}^{n}Y_j\begin{pmatrix}1\\ X_{ij}\end{pmatrix}K\left(\frac{||S_kX_{ij}||^2}{h_k^2}\right),$$

其中 $X_{ij}=X_j-X_i$.

(6) 利用第三步得到的 w_i 计算向量 $\hat{\gamma}_k=\left(\sum\limits_{j=1}^{n}w_i\right)^{-1}\sum\limits_{i=1}^{n}\widehat{\nabla m}_k(X_i)w_i$. 如果 $||\hat{\gamma}_k||>1$, 那么标准化 $\hat{\gamma}_k:=\hat{\gamma}_k/||\hat{\gamma}_k||$.

(7) 设置 $\rho_{k+1} = a_\rho \rho_k$, $h_{k+1} = \min\{a_h h_k, h_{\max}\}$. 如果 $\rho_{k+1} > \rho_{\min}$, 那么设置 $k = k+1$, 并继续步骤 2, 否则结束.

用 $k(n)$ 表示迭代的总数, 最后迭代的估计量 $\hat{\gamma} = \hat{\gamma}_{k(n)}$ 可以用来构造 β_0 的估计量: $\hat{\beta} = \hat{\gamma}/||\hat{\gamma}||$.

值得提到的是, 在上述算法中迭代的次数总是有限的, 因此在迭代中估计量的收敛性没有问题.

2. 选择算法的参数

显然, 由上述方法得到的估计强烈地依赖于变换参数 h 和 ρ 的规则, 尤其它们在初始和最终迭代的值. 在迭代期间 h_k 从 h_1 到 $h_{\max}$ 递增, 而 ρ_k 从 $\rho_1 = 1$ 到 $\rho_{\min}$ 递减. 用这样一种方法选择值 h_1, 使得对每一个 (或几乎每一个) 点 X_i 都能很好地定义 $\widehat{\nabla m}(X_i)$. 一个必要 (且通常充分) 的条件是每一个球 $\{x : ||x - X_i|| \leqslant h_1\}$ 至少包含 $p+1$ 个设计点. 在最终迭代 $k(n)$ 下 h 和 ρ 的值可以通过极小化估计量 $\hat{\beta}$ 的风险而得到. 这就导致下列建议: 值 $h_{k(n)}$ 在最终迭代时将尽可能大, 即大约为 1; 相反, 值 $\rho_{\min}$ 将选择得尽可能小, 但仍假定在每一个 (或几乎每一个) 局部椭球体邻域 $U_k(X_i) = \{x : ||S_k(x - X_i|| \leqslant h_k\}$ 有足够的设计点. 我们提出下面的经验规则:

$$\begin{aligned} &h_1 = n^{-1/(4\vee p)}, \quad &&h_{\max} = 2\sqrt{p}, \quad &&a_h = \mathrm{e}^{1/2(4\vee p)}, \\ &\rho_1 = 1, &&\rho_{\min} = n^{-1/3}/h_{\max}, &&a_\rho = \mathrm{e}^{-1/6}. \end{aligned} \tag{3.4.9}$$

假如 $h_{k(n)} \approx h_{\max}$, 显然由 (3.4.9) 式可得到迭代数目 $k(n) \approx \log_{a_\rho}(\rho_1/\rho_{\min}) = 2\log n$.

Hristache 等 (2001) 证明了估计量 $\hat{\beta}$ 是 $\sqrt{n}$ 相合的, 且 $\sqrt{n}(\hat{\beta} - \beta_0)$ 依分布收敛到正态向量.

3.5 切片逆回归估计

切片逆回归 (sliced inverse regression, SIR) 是用来估计指标向量的投影方向的一个联系自由的回归方法. 所谓 "联系自由" 就是不要求指明联系函数的具体形式. SIR 的一个主要优点是非常容易实施且不需要迭代计算.

3.5.1 多指标模型

本节考虑更一般的多指标模型

$$Y = G(\beta_1^{\mathrm{T}} X, \cdots, \beta_d^{\mathrm{T}} X, \varepsilon), \tag{3.5.1}$$

其中 X 是 p 维随机变量, 诸 β 是未知向量, 随机误差 ε 独立于 X, $G(\cdot)$ 是 $\mathbf{R}^{d+1}$ 上的任意未知函数.

模型 (3.5.1) 的含义是将 p 维协变量 X 投影到 d 维子空间 $(\beta_1^{\mathrm{T}}X,\cdots,\beta_d^{\mathrm{T}}X)$, 捕获所有需要知道的响应变量 Y. 当 d 较小时, 通过有效地估计诸 β 可以达到数据降维的目的. 为方便起见, 把诸 β 的任何线性组合看作有效降维方向, 并把由诸 β 生成的线性空间 $\mathcal{B}$ 看作有效降维空间. 本节主要考虑有效降维方向的估计, 不再考虑如何估计函数 $G(\cdot)$ 的问题. 直观地说, 估计出有效降维方向以后, 通常的光滑技术可以更成功地应用于 $G(\cdot)$ 的估计问题, 因为此时维数已降低.

本节估计有效降维方向的方法是基于逆回归的思想. 逆回归函数 $E(X|Y=y)$ 是 X 关于 Y 的回归. 它避开了所谓的 "维数灾祸" 问题, 其本质是用来处理一维到一维的回归问题, 而不是高维前向回归.

3.5.2 逆回归曲线

考虑逆回归曲线 $\zeta(y)=E(X|Y=y)$ 的轨迹. 这个曲线的中心位于 $E[E(X|Y)]=E(X)$. 中心化曲线 $\zeta(y)-E(X)$ 是 $\mathbf{R}^p$ 上的 p 维曲线. 对设计分布要求满足下列条件:

条件 3.5.1 对任何 $b\in\mathbf{R}^p$, 条件期望 $E(b^{\mathrm{T}}X|\beta_1^{\mathrm{T}}X,\cdots,\beta_d^{\mathrm{T}}X)$ 是 $\beta_1^{\mathrm{T}}X,\cdots,\beta_d^{\mathrm{T}}X$ 的线性函数; 即对某些常数 $c_0,c_1,\cdots,c_d$, 有

$$E(b^{\mathrm{T}}X|\beta_1^{\mathrm{T}}X,\cdots,\beta_d^{\mathrm{T}}X)=c_0+c_1\beta_1^{\mathrm{T}}X+\cdots+c_d\beta_d^{\mathrm{T}}X.$$

当 X 的分布为椭球对称 (正态) 分布时, 条件 3.5.1 满足. 条件 3.5.1 好像是强加给 X 的分布的一个严格的要求. 一个蕴涵是: 在数据收集阶段应当设计好试验, 以致 X 的分布不很明显地违反椭球对称, 除非响应面的函数形式已知. Li(1991) 证明了下列结果.

定理 3.5.1 在条件 3.5.1 下, 中心化的回归曲线 $\zeta(y)-E(X)$ 包含在由 $\beta_k^{\mathrm{T}}\Sigma_X$ $(k=1,\cdots,d)$ 张成的子空间内, 其中 Σ_X 是 X 的协方差阵.

根据定理 3.5.1, 可以得到

$$\beta_k\propto\Sigma_X^{-1}[\zeta(y)-E(X)],\quad k=1,\cdots,d.$$

使用上式右端可以寻找 β_0 的方向.

为方便起见, 考虑 X 的标准化形式: $Z=\Sigma_X^{-1/2}[X-E(X)]$. 此时, 模型 (3.5.1) 可以重写为

$$Y=G(\eta_1^{\mathrm{T}}Z,\cdots,\eta_d^{\mathrm{T}}Z,\varepsilon),\tag{3.5.2}$$

其中 $\eta_k=\Sigma_X^{1/2}\beta_k(k=1,\cdots,d)$. 我们将在由诸 η_k 的标准有效降维方向生成的线性空间上命名任何向量.

推论 3.5.1 在模型 (3.5.2) 和条件 3.5.1 下, 标准化逆回归曲线 $E(Z|Y)$ 包含在由标准有效降维方向 $\eta_1,\cdots,\eta_d$ 生成的线性空间内.

这个推论的一个重要结果是协方差阵 $\text{cov}(E(Z|Y))$ 在任何正交于 η_0 的方向上是退化的. 因此, 我们看到与 $\text{cov}(E(Z|Y))$ 的最大特征值对应的特征向量是标准有效降维方向. 这就导出了下一节的 SIR 算法.

推论 3.5.1 的一个有趣的扩展是: 当逆回归曲线 $E(Z|T)$ 轻微地违背条件 3.5.1 时, 确定偏离标准有效降维空间到什么程度的数量指标. 如果在标准有效降维空间的正交补上 $E(Z|T)$ 的投影是小的, 那么通过关于 $\text{cov}(E(Z|Y))$ 的主成分分析得到的方向仍接近标准有效降维方向. 这种情况类似 Brillinger(1977, 1983) 证明的结论: 在条件 3.5.1 下对单分量模型 β_1 的最小二乘估计具有相合性. 此外, 经验表明这个结果对条件 3.5.1 的违反是不敏感的. Li(1989) 给出了最小二乘的稳健问题的综合考虑和回归估计的一般使用.

条件协方差阵 $\text{cov}(Z|Y)$ 也为寻找标准有效降维方向提供了极有用的线索. 为看到这一点, 简单计算可知

$$E[\text{cov}(Z|Y)] = \text{cov}(Z) - \text{cov}(E(Z|Y)) = I - \text{cov}(E(Z|Y)).$$

因此, $E[\text{cov}(Z|Y)]$ 的特征值分解以后, 我们可以从联系最大特征值的特征向量中寻找标准有效降维方向. 此外, 估计 $E[\text{cov}(Z|Y)]$ 是不困难的, 具体做法将在下节介绍.

回归模型通常是通过将 (X,Y) 的联合密度函数分解为 $f(y|x)f(x)$ 而形成的. 这是回归的前向观察. 回归的逆向观察是将联合密度函数分解 $f(x|y)f(y)$ 而建模 $f(x|y)$. 逆形式的重要情形包括判别分析. SIR 从建模的逆向考虑其自身富有意义. 考虑将 Y 作为具有经验 Bayes 先验的参数的研究是有趣的.

3.5.3 运算步骤

假设 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 是来自模型 (3.5.1) 的独立同分布样本. 下面应用定理 3.5.1 来估计诸 β_k. 具体运算步骤如下:

(1) 通过一个仿射变换标准化 X_i, 得到 $\widetilde{X}_i = \Sigma_X^{-1/2}(X_i - \bar{X})(i=1,\cdots,n)$, 其中 $\bar{X}$ 和 Σ_X 分别是 X 的样本均值和样本协方差阵.

(2) 将 Y 的取值范围分成 H 片, 并记为 $I_1,\cdots,I_H$; 设 Y_i 落在片 h 内的比率为 $\hat{p}_h$, 即 $\hat{p}_h = n^{-1}\sum_{i=1}^{n}\delta_h(Y_i)$, 其中 $\delta_h(Y_i)$ 取值 0 或 1 依赖于是否 Y_i 落入第 h 片 I_h.

(3) 在每一片内, 计算 $\widetilde{X}_i$ 的样本均值, 并记为 $\hat{m}_h(h=1,\cdots,H)$, 即 $\hat{m}_h = (n\hat{p}_h)^{-1}\sum_{Y_i\in I_h}\widetilde{X}_i$.

(4) 计算加权协方差阵 $\hat{V}=\sum_{h=1}^{H}\hat{p}_h\hat{m}_h\hat{m}_h^{\mathrm{T}}$, 然后求 $\hat{V}$ 的特征值和特征向量.

(5) 设 d 个最大特征向量为 $\hat{\eta}_k(k=1,\cdots,d)$. 输出 $\hat{\beta}_k=\hat{\eta}_k^{\mathrm{T}}\Sigma_X^{-1/2}(k=1,\cdots,d)$.

步骤 (2) 和 (3) 对标准化逆回归曲线提供了一个粗糙的估计. 当然也可以用其他的非参数回归估计方法对它进行估计, 如核估计、最紧邻估计、光滑样条估计等. 这里我们主张使用这种切片的方法是由于它的操作简单.

在步骤 (4) 中, 对主成分分量的加权调整是为了处理在不同切片中可能有不等的样本量的情况. 前 d 个分量指出了最重要的子空间去跟踪标准逆回归曲线 $E(Z|Y=y)$. 最后, 步骤 (5) 经过尺度变换而得到诸 $\hat{\beta}_k$ 可以用来估计有效降维方向, 且有效降维空间 $\mathcal{B}$ 用 $\hat{\mathcal{B}}$ 估计, 其中空间 $\hat{\mathcal{B}}$ 由诸 $\hat{\beta}_k$ 产生.

把每个个体 X_i 变换为 $\widetilde{X}_i$ 并不是必需的. 在控制主成分分析之前作变换是为了节省计算时间. 设 $\hat{\Sigma}_1=\sum_{h=1}^{H}\hat{p}_h(\bar{X}_h-\bar{X})(\bar{X}_h-\bar{X})^{\mathrm{T}}$, 其中 $\bar{X}_h$ 表示在第 h 片上诸 X_i 的样本均值. 那么诸 $\hat{\beta}_k$ 恰好是 $\hat{\Sigma}_1$ 的特征值对应的特征向量, 这里 $\hat{\Sigma}_1$ 和 $\hat{\Sigma}_X$ 有类似作用. $\hat{\Sigma}_1$ 描述片间的变动程度.

需要提到的是: 每片的范围可以设置为等长, 也可以是可变的. 在具体应用中, 掌握的一个原则是每片的观察值尽可能相互接近.

另一个问题是片数的选择, 它可能影响估计的渐近方差. 这个问题的选择准则比非参数回归中光滑参数的选择准则少的多. 理论上, 在非参数回归中光滑参数的不恰当选择可能导致过低的收敛速度. 不论 H 如何选取, 使用上述算法得到的估计量仍然有 $\sqrt{n}$ 相合性. 数值模拟表明其结果也没有明显的差别. 参看文献 (Li, 1991).

当标准化 X 时, 把仿射变换建立在样本均值和样本方差阵基础上不是必需的. 这样做的目的是至少我们有可能剔除对稳健性有影响的设计点. 但这个问题可能是较少决定性的且相对容易处理, 因为是在我们的控制下处理设计点 (即使在观察研究中, 我们可以筛选出坏的设计点; 如果剩下的点的比例足够高, 我们仍有好的分析结果). 注意到仿射变换不是主要目的, 我们做 SIR 工作仅仅是需要有一个相合估计.

如果标准逆回归曲线落在一个标准有效降维特征子空间内, 那么 SIR 不能找回所有有效降维方向. 例如, 如果 $Y=g(\beta_1^{\mathrm{T}}X)+\varepsilon$, 对某个对称函数 $g(\cdot)$, 且 $\beta_1^{\mathrm{T}}X$ 也是关于 0 对称, 那么 $E(X|Y)=0$, 且 $\hat{\beta}_1$ 是 β_1 的一个低劣的估计. 为处理这种情况, 一个方法是探讨给定 Y 下 X 的高阶条件矩. 例如, 如果 X 是正态的, 那么对于任何正交于诸 $\beta_k^{\mathrm{T}}X$ 的方向 $b^{\mathrm{T}}X$, 我们看到 Y 改变时 $\mathrm{var}(b^{\mathrm{T}}X|Y)$ 仍是不变量. 特别, 对于标准化的 X, $\mathrm{cov}(X|Y\in I_h)$ 的异于 1 的特征值的特征向量在有效降维空间内. 因此, 如果对每片 h 在诸 $\widetilde{X}_i$ 的样本协方差 cov_h 上特征值分解被控制, 那

么可以用适当方式结合来自每片异于 1 的特征值的特征向量来估计 β_0. 关于结合方法的细节在研究中. 然而, 对每片分别控制特征值分解不一定必需. 例如, 人们可以把它看作用一个非负定矩阵分别拟合每个 $\text{cov}_h - I$ 的逼近问题.

Li(1991) 提出了一个统计量来阐述由 SIR 建立的方向是否真实的问题, 同时也证明了 SIR 带来有效降维方向的 $\sqrt{n}$ 相合估计. Hsing 和 Carroll(1992) 在适当的条件下证明了 SIR 估计的 $\sqrt{n}$ 相合性和渐近正态性. Duan 和 Li(1991) 研究了单指标模型中兴趣参数的 SIR 估计问题, 并证明了 SIR 估计的渐近正态性. Zhu 和 Fang(1996) 进一步研究了 SIR 的核估计的渐近性质. 这里不再一一赘述.

3.6 最小平均方差估计

最小平均方差估计方法由 Xia 等 (2002b) 提出. 他们考虑了多指标模型

$$Y = g(B_0^{\mathrm{T}} X) + \varepsilon, \tag{3.6.1}$$

其中 $g(\cdot)$ 是一个未知的光滑函数, $B_0 = (\beta_{01}, \cdots, \beta_{0D})$ 是 $p \times D$ 正交矩阵 ($B_0^{\mathrm{T}} B_0 = I_{D\times D}$), $D < p$, 且几乎处处有 $E(\varepsilon|X) = 0$.

3.6.1 有效降维方向的估计

用 d 记为作业维数, $1 \leqslant d \leqslant p$. 因此, 我们仅需要估计一束正交向量. 按照 Xia 等 (2002) 提出的方法, 方向 B_0 是极小化问题

$$\min_B \{E[Y - E(X|B^{\mathrm{T}} X)]^2\} \tag{3.6.2}$$

的解. 对任何正交矩阵 $B = (\beta_1, \cdots, \beta_d)$, 给定 $B^{\mathrm{T}} X$ 下的条件方差为

$$\sigma_B^2(B^{\mathrm{T}} X) = E\{[Y - E(X|B^{\mathrm{T}} X)]^2 | B^{\mathrm{T}} X\}.$$

由此推出

$$E[\sigma_B^2(B^{\mathrm{T}} X)] = E[Y - E(X|B^{\mathrm{T}} X)]^2.$$

因此, 关于 B 的极小化问题 (3.6.2) 等价于极小化

$$\sigma_B^2(B^{\mathrm{T}} X), \quad B^{\mathrm{T}} B = I. \tag{3.6.3}$$

记此估计为 $\hat{B}$, 并称它为最小平均方差估计 (minimum average variance estimation), 简记为 MAVE. 设 $\{(X_i, Y_i), 1 \leqslant i \leqslant n\}$ 是来自 (X, Y) 的 iid 样本. 记

$$g_B(u_1, \cdots, u_d) = E(Y|\beta_1^{\mathrm{T}} X = u_1, \cdots, \beta_d^{\mathrm{T}} X = u_d).$$

对任何给定的 X_0, $g_B(B^{\mathrm{T}}X_i)$ 在 X_0 处的局部线性展开为

$$g_B(B^{\mathrm{T}}X_i) \approx a + b^{\mathrm{T}}B^{\mathrm{T}}(X_i - X_0), \tag{3.6.4}$$

其中 $a = g_B(B^{\mathrm{T}}X_0)$, $b = (b_1, \cdots, b_d)^{\mathrm{T}}$,

$$b_k = \left.\frac{\partial g_B(u_1, \cdots, u_d)}{\partial u_k}\right|_{u_1=\beta_1^{\mathrm{T}}X_0, \cdots, u_d=\beta_d^{\mathrm{T}}X_0}, \quad k = 1, \cdots, d.$$

那么可得残差的近似表达:

$$Y_i - g_B(B^{\mathrm{T}}X_i) \approx Y_i - [a + b^{\mathrm{T}}B^{\mathrm{T}}(X_i - X_0)].$$

按照局部线性估计的思想, 并利用下面近似表达式可以估计 $\sigma_B^2(B^{\mathrm{T}}X_0)$,

$$\sum_{i=1}^{n}[Y_i - g_B(B^{\mathrm{T}}X_i)]^2 w_{i0} \approx \sum_{i=1}^{n}\{Y_i - [a + b^{\mathrm{T}}B^{\mathrm{T}}(X_i - X_0)]\}^2 w_{i0}, \tag{3.6.5}$$

其中 w_{i0} 是某个非负权, $\sum_{i=1}^{n} w_{i0} = 1$, 且以 $B^{\mathrm{T}}X_0$ 为中心. 在收缩有效降维方向中, 权 w_{i0} 起着关键作用. 这个问题将在后面详细讨论. 通常取

$$w_{i0} = K_h(B^{\mathrm{T}}(X_i - X_0)) \Big/ \sum_{j=1}^{n} K_h(B^{\mathrm{T}}(X_j - X_0)),$$

其中 $K_h(\cdot) = h^{-d}(\cdot/h)$, $K(\cdot)$ 是一个 d 元核函数, $h = h_n$ 为带宽. 为阐述方便起见, $K(\cdot)$ 在不同平面上表示不同的核函数. a 和 b 的估计量恰好是近似式 (3.6.5) 的极小值点. 因此, $\sigma_B^2(B^{\mathrm{T}}X_0)$ 的估计量恰好是表达式 (3.6.5) 的极小值, 即

$$\hat{\sigma}_B^2(B^{\mathrm{T}}X_0) = \min_{a,b}\left\{\sum_{i=1}^{n}\{Y_i - [a + b^{\mathrm{T}}B^{\mathrm{T}}(X_i - X_0)]\}^2 w_{i0}\right\}. \tag{3.6.6}$$

在一些宽松的条件下, 有 $\hat{\sigma}_B^2(B^{\mathrm{T}}X_0) - \sigma_B^2(B^{\mathrm{T}}X_0) = o_P(1)$. 以 (3.6.2)、(3.6.3) 和 (3.6.6) 式为基础, 通过求解下列极小化问题可以估计有效降维方向:

$$\begin{aligned}
&\min_{B:B^{\mathrm{T}}B=I}\left\{\sum_{j=1}^{n}\hat{\sigma}_B^2(B^{\mathrm{T}}X_j)\right\} \\
=&\min_{\substack{B:B^{\mathrm{T}}B=I\\ a_j, b_j, j=1,\cdots,n}}\left\{\sum_{j=1}^{n}\sum_{i=1}^{n}\{Y_i - [a_j + b_j^{\mathrm{T}}B^{\mathrm{T}}(X_i - X_j)]\}^2 w_{ij}\right\},
\end{aligned} \tag{3.6.7}$$

其中 $b_j = (b_{j1}, \cdots, b_{jd})^{\mathrm{T}}$. MAVE 方法或极小化问题 (3.6.7) 可以看作同时实施有效降维方向估计和非参数函数估计.

如果权依赖于 B, 极小化问题 (3.6.7) 的实施是非平凡的. 选择 (3.6.5) 式中的权 w_{i0} 使得 w_{i0} 的值是 X_i 和 X_0 之间的距离的函数. 下面给出 w_{i0} 的两个选择.

1. 多维核权

为简化问题 (3.6.7), 权 w_{i0} 的一个自然地选择是

$$w_{i0}=K_h(X_i-X_0)\Big/\sum_{j=1}^{n}K_h(X_j-X_0).$$

这种权可以用作估计的初始步骤. 给定 d, 我们利用极小化问题 (3.6.7) 得到一束方向 $\hat{B}$. 用 $S(\hat{B})$ 表示由 $\hat{B}$ 的列向量张成的子空间, 且 $S(B_0)$ 表示由 B_0 的列向量张成的空间. 用 $\{||(I-B_0B_0^{\mathrm{T}})\hat{B}||,d<D\}$ 和 $\{||(I-\hat{B}\hat{B}^{\mathrm{T}})B_0||,d>D\}$ 度量 $S(\hat{B})$ 和 $S(B)$ 之间的距离, 并记这个距离为 $m(\hat{B},B_0)$. 在一定的正则条件下, Xia 等 (2002) 证明了下列结果:

设 $h\to 0$, $nh^p/\log n\to\infty$. 如果 $d<D$, 则

$$m(\hat{B},B_0)=O_P(h^2+h^{-1}\delta_n^2);$$

如果 $d\geqslant D$, 则

$$m(\hat{B},B_0)=O_P(h^3+h^{-1}\delta_n^2),$$

其中 $\delta_n=(nh^p/\log n)^{-1/2}$.

假如维数被正确选择, $\hat{B}$ 的收敛速度是 $O_P(h_{\mathrm{opt}}^3\log n)$, 其中 h_{opt} 是极小化积分均方误差意义下回归函数估计的最优带宽. 这个速度比用其他方法获得的速度 $O_P(h_{\mathrm{opt}}^2)$ 更快. 这个更快的速度是由于极小化关于方向和联系函数的局部线性化的平均 (条件) 方差的缘故. 此外, 如果把这个思想扩展到高阶局部多项式光滑, B_0 的估计量的 $\sqrt{n}$ 相合性仍成立. 见文献 (Xia et al., 2002b).

2. 改良的核权

如果知道有效降维空间的维数小于 p, 那么我们可以在低维空间中搜寻有效降维方向, 由此, 减少高维的影响并提高估计的精度. 假设有 B_0 的初始估计 $\hat{B}$. 记

$$\widetilde{w}_{ij}=K_h(\hat{B}^{\mathrm{T}}(X_i-X_j))\Big/\sum_{l=1}^{n}K_h(\hat{B}^{\mathrm{T}}(X_l-X_j)), \tag{3.6.8}$$

用 $\widetilde{w}_{ij}$ 代替 (3.6.7) 式中的 w_{ij}, 并极小化问题 (3.6.7) 来重新估计 B_0. 为简化记号, 我们仍用 $\hat{B}$ 记这个新估计. 用最新的估计代替 (3.6.8) 式中的 $\hat{B}$. 重复这个步骤直到 $\hat{B}$ 收敛; 称这个极限为改良的 MAVE(RMAVE). 可以得到类似定理 3.5.1 的结果. 这里我们使用低维核和比 w_{ij} 中更小的带宽导致更快的收敛速度.

3.6.2 有效降维空间的维数

Li(1991) 提出了有效降维方向数目的确定方法. 该方法是基于由 SIR 强加给协变量 X 的假定. 下面提出一个平行的方法. 众所周知, 交叉验证方法是对模型复杂性的惩罚, 参看文献 (Stone, 1974). 下面将 Cheng 和 Tong(1992) 的交叉验证法推广到解上述问题.

设 $\beta_1,\cdots,\beta_D$ 是有效降维方向, 即 $Y=g(\beta_1^{\mathrm{T}}X,\cdots,\beta_D^{\mathrm{T}}X)+\varepsilon$, 且几乎处处有 $E(\varepsilon|X)=0$. 如果 $D<p$, 我们可以把方向的数目扩大到 p, 记作 $\{\beta_1,\cdots,\beta_D,\cdots,\beta_p\}$, 使得它们相互正交. 现在, 这个问题变为在 $\{\beta_1^{\mathrm{T}}X,\cdots,\beta_p^{\mathrm{T}}X\}$ 中的协变量的选择问题. 然而, 因为 $\beta_1,\cdots,\beta_p$ 是未知的, 我们必须用它们的估计量 $\hat{\beta}_k$ 代替诸 β_k. 正如已经证明的结论: $\hat{\beta}_k$ 的收敛速度比非参数联系函数的收敛速度更快. 这个代替是合乎情理的. 记

$$\hat{a}_{d0,j}=\sum_{i=1,i\neq j}^{n}K_{h_d}^{(i,j)}Y_i\Big/\sum_{i=1,i\neq j}K_{h_d}^{(i,j)},$$

其中 $K_{h_d}^{(i,j)}=K_{h_d}(\hat{\beta}_1^{\mathrm{T}}(X_i-X_j),\cdots,\hat{\beta}_p^{\mathrm{T}}(X_i-X_j))$. 这里我们对带宽使用下标 d 是为了强调它依赖于作业维数 d. 记

$$\mathrm{CV}(d)=\frac{1}{n}\sum_{j=1}^{n}(Y_j-\hat{a}_{d0,j})^2,\quad d=1,\cdots,p.$$

极小化 CV(d) 即可得到估计有效降维空间的维数, 即

$$\hat{d}=\arg\min_{0\leqslant d\leqslant p}\{\mathrm{CV}(d)\}.$$

在适当的条件下, Xia 等 (2002) 证明了

$$\hat{d}\xrightarrow{P}D.$$

3.6.3 带宽和算法

MAVE 方法的一个重要特征是不需要对有效降维方向估计中的联系函数估计做欠光滑以达到更高的收敛速度. 因此, 可以使用在积分均方误差意义下的最优带宽, 在实际中, 通常推荐使用变带宽, 即

$$K_{\mathbf{h}}(u_1,\cdots,u_d)=K(u_1/h_{(1)},\cdots,u_d/h_{(d)})\Big/\prod_{k=1}^{n}h_{(k)},$$

其中 $\boldsymbol{h}=(h_{(1)},\cdots,h_{(d)})$, d 是 K 的维数. 有许多方法可以得到这样一个带宽 $\boldsymbol{h}$(Fan and Gijbels, 1996). 我们的搜索程序如下.

步骤 1 (方向). 对每个 $d, 1\leqslant d\leqslant p$, 我们搜索 d 个方向如下.

(a) 初始值: 使用多维核并通过极小化问题 (3.6.7) 得到一个合适的有效降维方向的初始估计 $\hat{\beta}_1, \cdots, \hat{\beta}_d$.

(b) 精确估计: 记 $\hat{B} = (\hat{\beta}_1, \cdots, \hat{\beta}_d)$ 构成最后的估计. 因此, 使用权 (3.6.8) 得到精确的估计. 使用改良的核权并利用表达式 (3.6.7) 更新估计量. 继续这个步骤直到收敛. 使用有效降维方向的最后估计可以得到 $\mathrm{CV}(d)$ 的值.

步骤 2 (维数和输出结果). 比较 $\mathrm{CV}(d)$, $0 \leqslant d \leqslant p$. $\mathrm{CV}(d)$ 的最小值是估计的方向. 步骤 1(b) 中 B 的相应的估计量给出了估计的有效降维方向.

设 $\hat{B}_a$ 和 $\hat{B}_b$ 是步骤 1(b) 中 B 的两个相邻迭代的估计. 步骤 1(b) 的一个建议的停止规则是当几个相邻迭代的距离 $m(\hat{B}_a, \hat{B}_b)$ 每个小于前一步的公差时停止计算.

下面我们描述实施极小化问题 (3.6.7) 的一个方法. 对任何 d, 记 $B = (\beta_1, \cdots, \beta_d)$ 是步骤 1(a) 中的初始值 (令 $\beta_1 = \cdots = \beta_d = 0$). $B_{l,k} = (\beta_1, \cdots, \beta_{k-1})$, $B_{r,k} = (\beta_{k+1}, \cdots, \beta_d)$, $k = 1, \cdots, d$. 在约束 $B_{l,k}^{\mathrm{T}}\beta = 0$ 和 $B_{r,k}^{\mathrm{T}}\beta = 0$ 下, 极小化

$$S_{n,k} = \sum_{j=1}^{n}\sum_{i=1}^{n}\left\{Y_i - a_j - (X_i - X_j)^{\mathrm{T}}(B_{l,k}, \beta, B_{r,k})(c_j^{\mathrm{T}}, d_j^{\mathrm{T}}, e_j^{\mathrm{T}})^{\mathrm{T}}\right\}^2 w_{ij},$$

其中 c_j 为 $(k-1)\times 1$ 向量, d_j 为标量, e_j 为 $(d-k)\times 1$ 向量. 这是一个典型的约束正交规划问题. 记

$$C_j = \sum_{i=1}^{n} w_{ij}(X_i - X_j),$$

$$D_j = \sum_{i=1}^{n} w_{ij}(X_i - X_j)(X_i - X_j)^{\mathrm{T}},$$

$$E_j = \sum_{i=1}^{n} w_{ij} Y_i,$$

$$F_j = \sum_{i=1}^{n} w_{ij}(X_i - X_j) Y_i.$$

利用给定的 β, 可以得到极小化 $S_{n,k}$ 的 (a_j, c_j, d_j, e_j), 即

$$\begin{pmatrix} a_j \\ c_j \\ d_j \\ e_j \end{pmatrix} = \begin{pmatrix} 1 & C_j^{\mathrm{T}}(B_{l,k}, \beta, B_{r,k}) \\ (B_{l,k}, \beta, B_{r,k})^{\mathrm{T}} C_j & (B_{l,k}, \beta, B_{r,k})^{\mathrm{T}} D_j (B_{l,k}, \beta, B_{r,k}) \end{pmatrix}^{-1} \times \begin{pmatrix} E_j \\ (B_{l,k}, \beta, B_{r,k})^{\mathrm{T}} F_j \end{pmatrix},$$

$j = 1, \cdots, n$. 如果给定 a_j, c_j, d_j, e_j, 可以得到极小化 $S_{n,k}$ 的 β, 即

$$\begin{pmatrix} \beta \\ \lambda \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} d_j^2 D_j & \widetilde{B}_k \\ \widetilde{B}_k^{\mathrm{T}} & 0 \end{pmatrix}^{+} \begin{pmatrix} \sum_{j=1}^{n} d_j \{F_j - a_j C_j - D_j \widetilde{B}_k (c_j^{\mathrm{T}}, e_j^{\mathrm{T}})^{\mathrm{T}}\} \\ 0 \end{pmatrix},$$

其中 $\widetilde{B}_k = (B_{l,k}, B_{r,k})$, A^+ 表示矩阵 A 的 Moore-Penrose 逆, λ 是约束极小化的 Lagrange 乘子. 最后, 标准化 β.

3.7　惩罚样条估计

Yu 和 Ruppert(2002) 提出了惩罚样条估计方法来估计指标系数. 下面介绍这种方法. 我们可以用 B 样条估计未知单变量函数 $g(\cdot)$. 假设

$$g(u) = \delta_0 + \delta_1 u + \cdots + \delta_d u_d + \sum_{i=1}^{k} \delta_{d+i} (u - s_i)_+^d, \tag{3.7.1}$$

其中 $\{s_i\}_{i=1}^k$ 是样条节点. 定义样条系数向量 $\delta = (\delta_0, \delta_1, \cdots, \delta_{d+k})^{\mathrm{T}}$ 和样条基

$$B(u) = (1, u, \cdots, u^d, (u - s_1)_+^d, \cdots, (u - s_k)_+^d)^{\mathrm{T}}.$$

那么, (3.7.1) 式可化为 $g(u) = \delta^{\mathrm{T}} B(u)$. 记 $\theta = (\beta^{\mathrm{T}}, \delta^{\mathrm{T}})^{\mathrm{T}}$. 极小化

$$Q_{n,\lambda}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \{Y_i - \delta^{\mathrm{T}} B(\beta^{\mathrm{T}} X_i)\}^2 + \lambda \delta^{\mathrm{T}} D \delta$$

可得到 θ 的惩罚最小二乘估计, 记作 $\hat{\theta}$. 其中 $\lambda \geqslant 0$ 为惩罚参数, D 为确定的半正定矩阵. D 可以是对角矩阵其最后 k 个对角线上的元素为 1 其余全为 0. 这是惩罚 $g(\cdot)$ 的第 d 阶跳跃的平方和. Yu 和 Ruppert(2002) 讨论了节点和 λ 的选择问题. 他们建议选 5~10 个节点就足够了, 且节点将由估计的指标值的等距量代替; 通过极小化 GCV 得分来选择惩罚参数 λ.

类似于 Yu 和 Ruppert(2002,2004) 的证明, 可以证明估计量 $\hat{\theta}$ 的 $\sqrt{n}$ 相合性和渐近正态性.

3.8　Bayes 估计

近年来, Bayes 方法已被成功应用到许多其他非线性回归模型, 如可加半参数回归模型 (Biller, 2000), 半参数风险函数回归 (McKeague and Tighiouart, 2000), 具有测量误差的非参数回归 (Berry et al., 2002), 由参数联系函数确定的广义线性模型 (Ntzoufras et al., 2003), 单指标模型 (Antoniadis et al., 2004; Wang, 2009) 等. 本节主要阐述 Antoniadis 等 (2004) 提出的 Bayes 估计方法.

3.8.1 建模构想

本节仍考虑模型 (3.1.1), 其中随机误差 ε_i 相互独立且服从正态分布 $N(0,\sigma^2)$, σ^2 是未知参数. 用分层方式定义先验, 从 β_0 和 σ^2 开始, 然后就 B 样条逼近来指定给定 β_0 和 σ^2 下 $g(\cdot)$ 的条件分布. 随后选定各种超参数. 先验的具体确定如下.

(1) β_0 的先验是具有集结参数 $\lambda_{\text{prior}} \geqslant 0$ 和模式方向 $\beta_{0,\text{prior}}$ 的 Fisher-von Mises. 换句话说, β_0 的先验密度正比于 $\exp(\lambda_{\text{prior}}\beta_0^{\text{T}}\beta_{0,\text{prior}})$, β_0 属于 $\mathbf{R}^p$ 内的单位球面.

(2) σ^2 的先验是逆 gamma, 其密度正比于 $\sigma^{-2(A+1)}\exp(-B^{-1}\sigma^{-2})$, $\sigma^2>0$. 这里 $A>0$ 和 $B>0$ 是调整常数.

(3) 给定 β_0 和 σ^2, 用 B 样条基函数的线性组合表示联系函数:

$$g(u)=\sum_{j=1}^{K}\gamma_j B_j(u), \tag{3.8.1}$$

其中 B_j 是区间 $[a_{\beta_0},b_{\beta_0}]$ 上基于 $m+1$ 个等间隔节点的第 j 个 q 阶 B 样条基函数, $K=m+q$. 这里 $a_{\beta_0}=\min\{\beta_0^{\text{T}}X_1,\cdots,\beta_0^{\text{T}}X_n\}-\delta$, $b_{\beta_0}=\max\{\beta_0^{\text{T}}X_1,\cdots,\beta_0^{\text{T}}X_n\}+\delta$, δ 为某个正的常数. 对给定的 β_0 和 σ^2, 基系数有先验

$$\gamma=(\gamma_1,\cdots,\gamma_K)^{\text{T}}\sim N\left(\hat{\gamma}_{\beta_0},\sigma^2(\widetilde{X}_{\beta_0}^{\text{T}}\widetilde{X}_{\beta_0})^{-1}\right), \tag{3.8.2}$$

其中 $\widetilde{X}_{\beta_0}$ 是在用 "变换设计点"$\beta_0^{\text{T}}X_i(i=1,\cdots,n)$ 估计 $g(\cdot)$ 的 B 样条表示下的设计矩阵, 且 $\hat{\gamma}_{\beta_0}=(\widetilde{X}_{\beta_0}^{\text{T}}\widetilde{X}_{\beta_0}+\rho I)^{-1}\widetilde{X}_{\beta_0}Y$ 是 γ 的相应惩罚最小二乘 (或岭回归) 估计. 正整数 q 和 m 以及 $\delta>0$ 和 $\rho>0$ 是调整常数. B_j 和 $\widetilde{X}_{\beta_0}$ 的确切定义在下面给出.

由 (3.8.1) 式的 B 样条表示可把单指标模型写成

$$\boldsymbol{Y}=B(X\beta_0)\gamma+\varepsilon, \tag{3.8.3}$$

其中 B 是由 $B(u)_{ij}=B_j(u_i)$ 定义的矩阵值函数, $u=(u_1,\cdots,u_n)^{\text{T}}\in\mathbf{R}^n$, $\boldsymbol{Y}=(Y_1,\cdots,Y_n)^{\text{T}}$ 是响应向量, $X=(X_1,\cdots,X_n)^{\text{T}}$ 是原设计矩阵, $\varepsilon=(\varepsilon_1,\cdots,\varepsilon_n)^{\text{T}}$ 是观察误差向量. 在模型的这个形式中将设计矩阵记为 $\widetilde{X}_{\beta_0}=X\beta_0$. 注意这里使用的基函数 B_j 无疑在区间 $[a_{\beta_0},b_{\beta_0}]$ 上依赖于 β_0.

我们建议在范围 100~700 内选择 λ_{prior}, 在先验模式方向上随置信度而定. 模式方向 $\beta_{0,\text{prior}}$ 可以取作 β_0 的现有频率估计之一; 模拟表明使用 HJS 估计量 (Hristache et al., 2001) 可得到好的结果, 而尝试使用任意的 $\beta_{0,\text{prior}}$, 其结果很不理想. 当误差方差已知时, A 小 (比方说 0.001) 而 B 大 (比方说 100). 作为特殊的应用, 取 $q=2$, $m=15\sim 20$ 和 $\delta=0.001$, 使用 (Yu and Ruppert, 2002) 中的节点数目, 产生

了好的效果. 然而, 上述说明仅仅是一般的指导方针, 对任何具体的应用, 重要的是通过反复检查估计拟合的情况对 λ_{prior}, A, B 和 m 的各种不同选择进行反复试验来确定它们. 用经验 Bayes 程序能够处理这些调整参数的自动数据驱动选择, 这将在下节作进一步讨论.

3.8.2 调整

经过推导可得 β_0 的后验密度正比于

$$D(\beta_0) = (S(\beta_0) + 2/B)^{-(A+n/2)} \exp(\lambda_{\text{prior}} \beta_0^{\text{T}} \beta_{0,\text{prior}}), \tag{3.8.4}$$

其中

$$S(\beta_0) = \boldsymbol{Y}^{\text{T}}\boldsymbol{Y} - \boldsymbol{Y}^{\text{T}}\widetilde{X}_{\beta_0}\left\{\Sigma_0 + \frac{1}{2}\Sigma_\rho(I - \Sigma_0^{-1}\Sigma_\rho)\right\}\widetilde{X}_{\beta_0}^{\text{T}}\boldsymbol{Y}, \tag{3.8.5}$$

$\Sigma_\rho = (\widetilde{X}_{\beta_0}^{\text{T}}\widetilde{X}_{\beta_0} + \rho I)^{-1}$, Σ_0 是 Σ_ρ 将 ρ 置为 0 的形式. 对给定的 β_0, 使用通常的最小二乘估计 (即 $\rho = 0$ 时的 $\hat{\gamma}_{\beta_0}$) 代替 $\hat{\gamma}_{\beta_0}$, 并用 (3.8.5) 式的 $S(\beta_0)$ 降低误差平方和 (Kohn et al., 2001), 导致预测中严重的超光滑和高偏差; 这是几乎不出人意料的, 因为当时没有变量选择和调整特征参与到 $g(\cdot)$ 的先验.

下面介绍的方法是将惩罚最小二乘估计 $\hat{\gamma}_{\beta_0} = \arg\min_\gamma(||\boldsymbol{Y} - \widetilde{X}_{\beta_0}\gamma||^2 + \rho||\gamma||^2)$ 作为内部调整. 这里 $\rho \geqslant 0$ 被用来控制惩罚, I 是 $K \times K$ 单位矩阵. 更一般地, 也可以应用 Eilers 和 Marx(1996) 的 P 样条方法: 用一个半正定矩阵 Q 代替 I, 用 $\rho\gamma^{\text{T}}Q\gamma$ 代替 $\rho||\gamma||^2$.

所谓内部调整, 就是用通常的最小二乘估计作为 γ 的先验均值, 并通过用

$$S_E(\beta_0) = \boldsymbol{Y}^{\text{T}}\boldsymbol{Y} - \boldsymbol{Y}^{\text{T}}\widetilde{X}_{\beta_0}\Sigma_\rho\widetilde{X}_{\beta_0}^{\text{T}}\boldsymbol{Y}$$

代替 (3.8.5) 式的 $S(\beta_0)$ 来直接调整后验. 两个调整的方法在计算上都比 Smith 和 Kohn(1996) 的 Bernoulli 先验更便捷, 后者考虑了 B 样条基系数的变量选择.

光滑参数 ρ 是一个重要的超参数, 它需要由数据自动选择. 为此目的, 考虑一个适应于下节 MCMC 技术的 Bayes 程序. 在 MCMC 的每一步更新时利用广义交叉验证 (GCV) 选择一个最优的 ρ. 给定 β_0, 极小化

$$\text{GCV}(\rho) = \frac{n^{-1}||\boldsymbol{Y} - \widetilde{X}_{\beta_0}\gamma_{\beta_0}||^2}{[1 - n^{-1}\text{tr}(L(\rho))]^2},$$

其中 $L(\rho)$ 是相应于 ρ 的光滑样条算子, 即

$$L(\rho) = L_s(\rho) + [I - L_s(\rho)]\widetilde{X}_{\beta_0}\{\widetilde{X}_{\beta_0}^{\text{T}}[I - L_s(\rho)]\widetilde{X}_{\beta_0}\}^{-1}\widetilde{X}_{\beta_0}^{\text{T}}[I - L_s(\rho)],$$

$L_s(\rho) = \widetilde{X}_{\beta_0}\Sigma_\rho \widetilde{X}_{\beta_0}^{\mathrm{T}}$. 注意到 $L(\rho)$ 的迹可以用下述方法计算: 仅需要估计 $\dim(\beta_0)\times\dim(\beta_0)$ 的矩阵的迹. 用 β_0 的现行估计值代替 β_0 即可进行 $\mathrm{GCV}(\rho)$ 的极小化.

当这个程序在 MCMC 的每一步重复时, 采样器收敛到 β_0 的经验 Bayes 后验分布, 见文献 (Casella, 2001). 然而, 在 MCMC 的每一步都要修正 B 样条基 (节点), 这将带来计算上的麻烦. 我们已经注意到 ρ 主要用来控制估计联系函数 $g(\cdot)$ 的光滑度, 但相对来说对指标系数 β_0 的估计影响不大. 因此, 恰好在用 HJS 估计代替未知 β_0 的 MCMC 运行的开始来极小化 $\mathrm{GCV}(\rho)$ 是合理的.

3.8.3 MCMC 算法

我们使用随机步长 Metropolis 算法从调整的后验中抽样. 建议分布 (proposal distribution) 取为 Fisher-von Mises, 其集结参数 $\lambda_{\mathrm{prop}} \geqslant 0$ 和模式向量由 $\beta_0 = \xi$ 的现行值给定. Ulrich(1984) 的算法被用来产生来自 Fisher-von Mises 分布的样本. 需要校准集结参数 λ_{prop} 以确保采样器能充分探测状态空间. 通过试验发现 $\lambda_{\mathrm{prop}} = 1000$ 时对小和中等的数据集产生好的结果.

Antoniadis 等 (2004) 证明了下列结论: 对 β_0 的调整后验密度, 随机步长 Metropolis 采样器是几何各态历经的. 因此, 在实际中我们理应得到好的结果. 尤其, 各态历经的平均数迅速收敛, 中心极限定理成立, 用分批平均值的方法估计 Monte-Carlo 误差的标准技术是可利用的.

采样器产生一列单位模的指标向量值. 我们用这些值的均值或逐分量中位数来估计 β_0, 并重新标准化得到单位模向量. 给定这样一个估计 $\hat{\beta}$, 取 $\widetilde{X}_{\hat{\beta}}\hat{\gamma}_{\hat{\beta}}$ 作为在变换设计点上对响应向量的估计拟合.

基于可逆跳跃 MCMC 技术 (Green, 1995), Wang (2009) 开发了规范的程序来处理单指标模型, 包括 Bayes 变量选择, 指标系数估计和联系函数拟合. 他使用多项式样条逼近未知的联系函数, 且把节点的数目和位置看做随机变量来处理. 引入二分量随机变量来指示模型中是否包括预测变量. 当没有可利用的先验信息时, 选择均匀球面分布作为指标系数的先验分布. 通过指定样条系数和误差方差的一个共轭逆 Gamma 先验, 得到了其余未知量的边缘后验, 且把边缘先验看作目标分布以获得快速算法. 为抽取指标系数的条件后验分布的样本, 他设计了一个新的随机游动 Metropolis 采样器来实现抽样.

3.9 分位数回归估计

本节讨论单指标分位数回归模型的估计问题. 这个模型具有形式

$$Y = g(\beta_0^{\mathrm{T}}X) + \varepsilon, \quad E[\varphi(\varepsilon)|X] = 0, \quad \text{a.s.}, \tag{3.9.1}$$

其中 X 是 p 维协变量向量, Y 是实值响应变量, $g(\cdot)$ 是未知的单变量联系函数, $\varphi(\cdot)$ 是某个已知函数, β_0 是指标系数. 为了模型的可识别, 假定 β_0 满足 $||\beta_0||=1$ 且第一个分量是正的. 在模型 (3.9.1) 中, $g(\cdot)$ 实际上应是 $g_\tau(\cdot)$, β_0 应是 $\beta_{0,\tau}$, 二者是唯一确定的, 这里 $\tau\in(0,1)$ 为给定的分位数. 为了记号方便, 我们省略了下标 τ.

本节的主要目的是 β_0 的估计. 真参数 β_0 是下列极小化问题的解:

$$\beta_0=\arg\min_{\beta}E[\rho_\tau(Y-g(\beta_0^{\mathrm{T}}X))],\quad ||\beta||=1,\quad \beta_1>0,$$

其中损失函数 $\rho_\tau(u)=|u|+(2\tau-1)u$. 上式右端是平均损失, 它可被等价地写作

$$E[\rho_\tau(Y-g(\beta_0^{\mathrm{T}}X))]=E\{E[\rho_\tau(Y-g(\beta_0^{\mathrm{T}}X))|\beta^{\mathrm{T}}X]\},\tag{3.9.2}$$

其中 $E[\rho_\tau(Y-g(\beta_0^{\mathrm{T}}X))|\beta^{\mathrm{T}}X]$ 是条件平均损失, $g(\cdot)$ 是第 τ 个条件分位数函数.

假设 $\{(X_i,Y_i),1\leqslant i\leqslant n\}$ 是来自模型 (3.9.1) 的 iid 样本. 根据上面提出的思想并利用局部线性方法可得到 β_0 的估计, 即

$$\hat{\beta}=\arg\min_{\beta}\min_{(a_j,b_j)}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_\tau(Y_i-a_j-b_j\beta^{\mathrm{T}}X_{ij})K(\beta^{\mathrm{T}}X_{ij}/h),\tag{3.9.3}$$

其中 $X_{ij}=X_i-X_j$, $K(\cdot)$ 是核函数, h 是带宽.

通过迭代可以实现 (3.9.3) 式的两步极小化. 具体算法步骤如下.

步骤 0 (初始化步). 取平均导数估计 (Chaudhuri et al., 1997) 作为 β_0 的一个初始估计, 并将其标准化, 记为 $\hat{\beta}$.

步骤 1. 极小化

$$\min_{(a_j,b_j)}\sum_{i=1}^{n}\rho_\tau(Y_i-a_j-b_j\hat{\beta}^{\mathrm{T}}X_{ij})K(\hat{\beta}^{\mathrm{T}}X_{ij}/h),\tag{3.9.4}$$

得到 (a_j,b_j) 的估计 $(\hat{a}_j,\hat{b}_j)$, 其中 $j=1,\cdots,n$. 这里取 h 的一个估计的带宽, 使得 β_0 的估计达到最优.

步骤 2. 更新 β_0 的估计, 即极小化

$$\min_{\beta}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_\tau(Y_i-\hat{a}_j-\hat{b}_j\beta^{\mathrm{T}}X_{ij})K(\hat{\beta}^{\mathrm{T}}X_{ij}/h),\tag{3.9.5}$$

得到 β_0 的一个估计, 并将其标准化作为更新估计, 不妨仍记为 $\hat{\beta}$.

步骤 3. 重复步骤 1 和 2 直到收敛. 将得到的值再标准化即可作为 β_0 的估计值.

步骤 4. 利用第三步得到的估计值 $\hat{\beta}$ 可得到 $g(u)$ 的最终估计 $\hat{g}(u)=\hat{a}$, 其中 $(\hat{a},\hat{b})$ 通过极小化

$$\sum_{i=1}^{n}\rho_\tau(Y_i-a-b(\hat{\beta}^{\mathrm{T}}X_i-u))K((\hat{\beta}^{\mathrm{T}}X_i-u)/h) \tag{3.9.6}$$

而得到. 在这最后一步中, 取 h 的一个估计的带宽, 使得当 β_0 已知时 $g(\cdot)$ 的估计达到最优.

为了便于计算, 把 (3.9.5) 式改写成形式

$$\min_{\beta}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_\tau(Y_{ij}^{\hat{\beta}}-\beta^{\mathrm{T}}X_{ij}^{\hat{\beta}}), \tag{3.9.7}$$

其中 $Y_{ij}^{\hat{\beta}}=(Y_i-\hat{a}_j)K^{1/2}(\hat{\beta}^{\mathrm{T}}X_{ij}/h)$, $X_{ij}^{\hat{\beta}}=\hat{b}_jX_{ij}K^{1/2}(\hat{\beta}^{\mathrm{T}}X_{ij}/h)$. 因此, (3.9.4) 和 (3.9.7) 都是简单的线性分位回归的极小化问题, 可以利用现有的有效算法, 见文献 (Koenker, 2005).

在初始化步中, 我们仅对初始值按方式 $\mathrm{sign}_1\beta/||\beta||$ 进行标准化, 其中 sign_1 是 β 的第一个分量的符号. 来自平均导数估计的初始值有优良的性质, 且已经证明它具有 $\sqrt{n}$ 相合性. 在第二步中对 β_0 的一个估计做标准化也是必要的, 因为由 (3.9.6) 得到的值其模不一定等于 1.

带宽选择在局部核光滑中总是至关重要的, 因为它控制估计函数的弯曲度. 人们可以使用 CV 和 GCV 方法选择最优带宽, 这两种带宽选择方法的计算量都很大. Yu 和 Jones(1998) 在适当的假定下获得了渐近最优带宽. 注意到 (3.9.10) 中 $\hat{g}(u)$ 的渐近方差和 Yu 和 Jones(1998) 给出的表达式的相似性, 并按照同样的论证可以得到下列的拇指规则带宽 h_τ:

$$h_\tau=h_m[\tau(1-\tau)/\phi^2(\Phi^{-1}(\tau))]^{1/5}, \tag{3.9.8}$$

其中 $\phi(\cdot)$ 和 $\Phi(\cdot)$ 分别是标准正态分布的概率密度函数和分布函数,

$$h_m=\left\{\frac{\nu_0\mathrm{var}(Y|\beta_0^{\mathrm{T}}X=u)}{ng''(u)f(u)\kappa_2^2}\right\}^{1/5}.$$

这里 $f(u)$ 是 $\beta_0^{\mathrm{T}}X$ 的密度函数, $\kappa_2=\int u^2K(u)\mathrm{d}u$, $\nu_0=\int K^2(u)\mathrm{d}u$. 利用一个仅包含 τ 的乘数并与使用在均值回归中的 h_m 相结合, 使得 h_τ 具有了好的特性. 因为有许多 h_m 的现成算法, h_τ 当然也是可以利用的.

近似公式 (3.9.8) 提供了一个便于计算的方法来计算分位回归中难以得到的最优带宽. 然而, 这个近似公式是基于几个关键的假定. 明显地, 条件中位数和条件

均值的曲率是类似的, 响应变量的条件密度可以用正态密度逼近. 关于逼近的详细资料可从文献 (Yu and Jones, 1998; Yu and Lu, 2004) 中找到.

在适当的条件下, Wu 等 (2010) 证明了估计量 $\hat{g}(u)$ 和 $\hat{\beta}$ 的渐近正态. 即有下列两个结果:

$$\sqrt{nh}[\hat{g}(u) - g(u) - b(u)] \xrightarrow{D} N(0, \gamma^2(u)), \tag{3.9.9}$$

其中 $b(u) = \frac{1}{2}h^2\kappa_2 g''(u)$, $\gamma^2(u) = \tau(1-\tau)\nu_0[f(u)]^{-1}[f_Y(g(u)|u)]^{-2}$, $f_Y(\cdot|u)$ 是给定 u 下 Y 的条件密度;

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, \tau(1-\tau)C_1^- C_0 C_1^-), \tag{3.9.10}$$

其中

$$C_0 = E\left\{g'(\beta_0^{\mathrm{T}}X)^2[X - E(X|\beta_0^{\mathrm{T}}X)][X - E(X|\beta_0^{\mathrm{T}}X)]^{\mathrm{T}}\right\},$$

$$C_1 = E\left\{f_Y(g(\beta_0^{\mathrm{T}}X))g'(\beta_0^{\mathrm{T}}X)^2[X - E(X|\beta_0^{\mathrm{T}}X)][X - E(X|\beta_0^{\mathrm{T}}X)]^{\mathrm{T}}\right\},$$

其中 C_1^- 表示矩阵 C_1 的广义逆.

3.10 估计方程估计

本节利用估计方程估计方法构造模型 (3.1.1) 中参数分量的有效估计, 讨论所提出的估计量的渐近性质并给出详细证明. 本节的估计方法和证明思路来自文献 (Wang et al., 2010a).

3.10.1 估计方法

本节在同方差下考虑模型 (3.1.1), 即假定 $\mathrm{var}(\varepsilon|X) = \sigma^2 > 0$. 利用广义估计方程给出模型 (3.1.1) 中指标系数的一个新估计. 该估计比前面几种方法得到的估计有更小的渐近方差, 因此它是一个更有效的估计.

设 $\mathcal{B} = \{\beta \in \mathbf{R}^p : \|\beta\| = 1$, 且它的第一个非零元素是正的$\}$. 那么 β_0 是 $\mathcal{B}$ 的内点. 因此, 我们仅在 $\mathcal{B}$ 上搜寻 β_0. 如果 $g(\cdot)$ 已知, 那么 β_0 的最小二乘估计是目标函数

$$M(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - g(\beta^{\mathrm{T}}X_i)\right]^2, \quad \beta \in \mathcal{B} \tag{3.10.1}$$

的极小值点. 如果使用 Newton 算法找 $M(\beta)$ 的极小值, 需要计算 $M(\beta)$ 关于 β 的导数. 然而, $\|\beta\| = 1$ 意味着 β 是单位球面上的点, $g(\beta^{\mathrm{T}}X)$ 在边界点上不存在导数. 为此, 我们建议将 β 再参数化, 以便在 Euclid 空间 $\mathbf{R}^{p-1}$ 的一个约束区域上搜寻方向 β_0. 可以使用通常的关于 β_0 的 "去一分量" 方法; 见 (Yu and Ruppert, 2002;

Chang et al., 2010; Wang et al., 2010a). 不失一般性, 假定 β_0 的第 r 个分量是正的. 对 $\beta=(\beta_1,\cdots,\beta_p)^{\mathrm{T}}$, 用 $\beta^{(r)}=(\beta_1,\cdots,\beta_{r-1},\beta_{r+1},\cdots,\beta_p)^{\mathrm{T}}$ 表示去掉 β 的第 r 个分量 β_r 后的 $p-1$ 维参数向量. 那么, 新参数 $\beta_0^{(r)}$ 一定满足 $||\beta_0^{(r)}||<1$, 即 $\beta_0^{(r)}$ 是单位球的内点. 因此, β 在 $\beta_0^{(r)}$ 的邻域内是无限可微的. 注意到

$$\beta=\beta(\beta^{(r)})=(\beta_1,\cdots,\beta_{r-1},(1-||\beta^{(r)}||^2)^{1/2},\beta_{r+1},\cdots,\beta_p)^{\mathrm{T}}, \tag{3.10.2}$$

可得到 β 关于 $\beta^{(r)}$Jacobian 矩阵:

$$J_{\beta^{(r)}}=\frac{\partial\beta}{\partial\beta^{(r)}}=(\gamma_1,\cdots,\gamma_p)^{\mathrm{T}}, \tag{3.10.3}$$

其中 γ_s $(1\leqslant s\leqslant p, s\neq r)$ 是第 $s-1(s>r)$ 个元素为 1 的 $p-1$ 维单位向量, $\gamma_r=-(1-||\beta^{(r)}||^2)^{-1/2}\beta^{(r)}$.

现在考虑极小化 (3.10.1) 式中的 $M(\beta)$ 的问题. 注意到 $M(\beta)=M(\beta(\beta^{(r)}))=\widetilde{M}(\beta^{(r)})$, 因此通过极小化 $\widetilde{M}(\beta^{(r)})$ 可以得到 $\beta_0^{(r)}$ 的一个估计, 然后再利用一个变换就可以得到 β_0 的估计. 该估计等价于方程组

$$\begin{cases}\displaystyle\sum_{i=1}^{n}[Y_i-g(\beta^{\mathrm{T}}X_i)]g'(\beta^{\mathrm{T}}X_i)J_{\beta^{(r)}}^{\mathrm{T}}X_i=0,\\ ||\beta||-1=0.\end{cases}$$

的解. 这意味着我们把一个约束最小二乘问题转化为无约束最小二乘问题. 由于 $g(\cdot)$ 及其导数 $g'(\cdot)$ 是未知的, 需要用其估计量代替它们. 下面使用 Fan 和 Gijbels(1996) 提出的局部线性拟合来估计 $g(\cdot)$ 和 $g'(\cdot)$. 对 u 的一个邻域内的点 U, 可以用一个线性函数局部地逼近 $g(U)$, 即

$$g(U)\approx g(u)+g'(u)(U-u)\equiv a_0+a_1(U-u).$$

对于固定的点 β_0, $g(u)$ 和 $g'(u)$ 的局部线性估计 $\hat{g}(u;\beta_0)$ 和 $\hat{g}'(u;\beta_0)$ 定义为加权平方和

$$\sum_{i=1}^{n}\{Y_i-[a_0+a_1(\beta_0^{\mathrm{T}}X_i-u)]\}^2K_h(\beta_0^{\mathrm{T}}X_i-u)$$

的极小值点, 其中 $K_h(\cdot)=h^{-1}K(\cdot/h)$, $K(\cdot)$ 是核函数, $h=h_n$ 是趋于 0 的带宽序列. 极小化上式可得 $\hat{g}(u;\beta_0)=\hat{a}_0$ 和 $\hat{g}'(u;\beta_0)=\hat{a}_1$. 周知, 估计量 $\hat{g}'$ 比 $\hat{g}$ 的收敛速度低. 如果 $\hat{g}'$ 与 $\hat{g}$ 使用同一个带宽, 就会导致 β_0 的估计量 $\hat{\beta}$ 的收敛速度比 $\sqrt{n}$ 更低, 从而影响 $\hat{\beta}$ 的渐近正态性. 因此, 为了理论证明的需要, 我们对 $\hat{g}'$ 使用另一个带宽 h_1(Wang et al., 2010a), 即用 h_1 代替 $\hat{g}'$ 中的 h. 由最小二乘理论可以推出

$$\hat{g}(u;\beta_0)=\sum_{i=1}^{n}W_{ni}(u,\beta_0)Y_i, \tag{3.10.4}$$

$$\hat{g}'(u;\beta_0)=\sum_{i=1}^{n}\widetilde{W}_{ni}(u,\beta_0)Y_i, \tag{3.10.5}$$

其中

$$W_{ni}\big(u;\beta_0\big)=\frac{V_{ni}\big(u;\beta_0,h\big)}{\sum_{j=1}^{n}V_{nj}\big(u;\beta_0,h\big)}, \tag{3.10.6}$$

$$\widetilde{W}_{ni}\big(u;\beta_0\big)=\frac{\widetilde{V}_{ni}\big(u;\beta_0,h_1\big)}{\sum_{j=1}^{n}V_{nj}\big(u;\beta_0,h_1\big)},$$

$$V_{ni}\big(u;\beta_0,h\big)=K_h\big(\beta_0^{\mathrm{T}}X_i-t\big)\big[S_{n,2}\big(u;\beta_0,h\big)-(\beta_0^{\mathrm{T}}X_i-t)S_{n,1}\big(u;\beta_0,h\big)\big],$$

$$\widetilde{V}_{ni}\big(u;\beta_0,h_1\big)=K_{h_1}\big(\beta_0^{\mathrm{T}}X_i-t\big)\big[(\beta_0^{\mathrm{T}}X_i-t)S_{n,0}\big(u;\beta_0,h_1\big)-S_{n,1}\big(u;\beta_0,h_1\big)\big],$$

$$S_{n,l}\big(u;\beta_0,h\big)=\frac{1}{n}\sum_{i=1}^{n}\big(\beta_0^{\mathrm{T}}X_i-t\big)^lK_h\big(\beta_0^{\mathrm{T}}X_i-t\big),\quad l=0,1,2.$$

利用估计量 $\hat{g}$ 和 $\hat{g}'$, 可以定义 β_0 的估计 $\hat{\beta}$ 为方程组

$$\begin{cases}\sum_{i=1}^{n}\big[Y_i-\hat{g}(\beta^{\mathrm{T}}X_i;\beta)\big]\hat{g}'(\beta^{\mathrm{T}}X_i;\beta)J_{\beta^{(r)}}^{\mathrm{T}}X_i=0,\\ ||\beta||-1=0\end{cases} \tag{3.10.7}$$

的解. 称 $\hat{\beta}$ 为 β_0 估计方程估计.

利用 $\hat{\beta}$, 定义 $g(u)$ 的最终估计量为 $\hat{g}^*(u)=\hat{g}(u,\hat{\beta})$; σ^2 的估计量为

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}[Y_i-\hat{g}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})]^2. \tag{3.10.8}$$

3.10.2　主要结果

为得到估计量的渐近性质, 我们首先给出下列条件.

(C3.10.1) 对 β_0 邻域内的点 β, $\beta^{\mathrm{T}}X$ 的概率密度函数 $f(u)$ 在 $\mathcal{U}$ 上是正的, 且满足 1 阶 Lipschitz 条件, 其中 $\mathcal{U}=\{u=\beta^{\mathrm{T}}x:x\in\mathcal{X}\}$, $\mathcal{X}$ 为 X 的紧支撑; 进一步, $\beta_0^{\mathrm{T}}X$ 具有正的且有界的概率密度函数.

(C3.10.2) (i) 函数 $g(u)$ 在 $\mathcal{U}$ 上具有有界连续的二阶导数;

(ii) $\mu(u)=E(X|\beta_0^{\mathrm{T}}X=u)$ 的每个分量在 $\mathcal{U}$ 上满足 1 阶 Lipschitz 条件.

(C3.10.3) $E(\varepsilon|X)=0$, $\sup\limits_{x}E(\varepsilon^4|X=x)<\infty$.

(C3.10.4) (i) 核 $K(u)$ 为 R 上有界、连续和对称的概率密度函数, 且满足

$$\int_{-\infty}^{\infty} u^2 K(u)\mathrm{d}u \neq 0, \quad \int_{-\infty}^{\infty} u^8 K(u)\mathrm{d}u < \infty;$$

(ii) $K(u)$ 在 $\mathbf{R}$ 满足 1 阶 Lipschitz 条件.

(C3.10.5) $nh^2/\log^2 n \to \infty$, $nh^4 \log n \to 0$; $nhh_1^3/\log^2 n \to \infty$, $nh_1^5 = O(1)$.

注 3.10.1 条件 (C3.10.1) 是用来约束 $\beta^{\mathrm{T}}X$ 的密度函数远离 0; 这就确保了 $\hat{g}(u;\beta)$ 和 $\hat{g}'(u;\beta)$ 的分母在 $\mathcal{U}$ 上以高的速度远离 0. 条件 (C3.10.2) 是通常的光滑条件. 条件 (C3.10.3)~(C3.10.5) 是研究估计量的渐近性质所必需的条件.

记 $\mathcal{B}_n^* = \{\beta \in \mathcal{B} : ||\beta - \beta_0|| \leqslant cn^{-1/2}\}$, $c > 0$ 为常数. 在下列定理中, 我们需要约束 β 在 $\mathcal{B}_n^*$ 内. 这个约束的动机是出于下列事实: 因为我们期望 $\hat{\beta}$ 是 $\sqrt{n}$ 相合的, 所以需要寻找方程组 (3.10.7) 的一个解, 它与 β_0 的距离的阶为 $n^{-1/2}$. 从定理 3.10.2 可以看出 $||\hat{\beta} - \beta_0|| = O_P(n^{-1/2})$, 因此 $\hat{\beta} \in \mathcal{B}_n$. 故把 β 约束在圆锥体内并没有排除任何感兴趣的极小值, 也没有失去一般性. Härdle 等 (1993) 也给出了类似的约束. 使用这个约束仅仅是为了理论证明的需要. 在实际中, 我们仅需要假定 $M(\beta)$ 在 $\mathcal{B}$ 上有唯一的极小值点即可.

首先给出估计量 $\hat{g}(u;\beta)$ 和 $\hat{g}'(u;\beta)$ 的一致收敛速度.

定理 3.10.1 设条件 (C3.10.1)、(C3.10.2)(i)、(C3.10.3) 和 (C3.10.4) 成立, 则

$$\sup_{u\in\mathcal{U},\beta\in\mathcal{B}_n^*} |\hat{g}(u;\beta) - g(u)| = O\left(\left\{\frac{\log n}{nh}\right\}^{1/2} + h^2\right), \quad \text{a.s.},$$

$$\sup_{u\in\mathcal{U},\beta\in\mathcal{B}_n^*} |\hat{g}'(u;\beta) - g'(u)| = O\left(\left\{\frac{\log n}{nh_1^3}\right\}^{1/2} + h_1\right), \quad \text{a.s.}$$

下列结果给出了估计量 $\hat{\beta}$ 的渐近正态性.

定理 3.10.2 设条件 (C3.10.1)~(C3.10.5) 成立, 则

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N\left(0, \sigma^2 J_{\beta_0^{(r)}} (J_{\beta_0^{(r)}}^{\mathrm{T}} D J_{\beta_0^{(r)}})^{-} J_{\beta_0^{(r)}}^{\mathrm{T}}\right).$$

其中记号 A^- 表示矩阵 A 的广义逆,

$$D = E\{[X - E(X|\beta_0^{\mathrm{T}}X)][X - E(X|\beta_0^{\mathrm{T}}X)]^{\mathrm{T}} g'(\beta_0^{\mathrm{T}}X)^2\}.$$

从定理 3.10.2 可得到关于 $\hat{\beta}$ 和 β_0 之间绝对内积一个渐近结果, 它可以用来研究充分降维问题. Cook(1998, 2007) 给出了更详细的讨论.

推论 3.10.1 在定理 3.10.2 的条件下, 有

$$|\hat{\beta}^{\mathrm{T}}\beta_0| - 1 = O_P(n^{-1/2}),$$

其中 $|\hat{\beta}^{\mathrm{T}}\beta_0|$ 表示 $\hat{\beta}$ 与 β_0 的内积的绝对值.

从 Härdle 等 (1993) 给出的定理可以看出: β_0 的估计量有渐近方差 $\sigma^2 D^-$. 注意到 D 有无穷多个广义逆矩阵, 但与 $J_{\beta_0^{(r)}}$ 有联系的广义逆只有一个. 下面的定理表明: 定理 3.10.2 的方差–协方差阵小于 $\sigma^2 D^-$, 这意味着 $D^- - J_{\beta_0^{(r)}}(J_{\beta_0^{(r)}}^{\mathrm{T}} D J_{\beta_0^{(r)}})^- J_{\beta_0^{(r)}}^{\mathrm{T}}$ 是半正定矩阵. 我们使用通常的记号: 对两个半正定矩阵 A_1 和 A_2, $A_1 \geqslant A_2$ 表示 $A_1 - A_2$ 是半正定矩阵. 为表达方便起见, 不妨假定 β_0 的第一个分量大于 0, 即 $r=1$. 把矩阵 D 作如下分块:

$$D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix},$$

其中 D_{11} 是正的常数, D_{12} 是 $1\times(p-1)$ 矩阵, D_{12} 是 $(p-1)\times 1$ 矩阵, D_{22} 是 $(p-1)\times(p-1)$ 非负定矩阵. Wang 等 (2010a) 和 Cui 等 (2011) 给出了下列结果.

定理 3.10.3　在定理 3.10.2 的条件下, 有

(i) 如果矩阵 $J_{\beta_0^{(r)}}^{\mathrm{T}} D J_{\beta_0^{(r)}}$ 是可逆的, 则 D 有一个形如 $J_{\beta_0^{(r)}}(J_{\beta_0^{(r)}}^{\mathrm{T}} D J_{\beta_0^{(r)}})^{-1} J_{\beta_0^{(r)}}^{\mathrm{T}}$ 的广义逆;

(ii) $J_{\beta_0^{(1)}}(J_{\beta_0^{(1)}}^{\mathrm{T}} D J_{\beta_0^{(1)}})^- J_{\beta_0^{(1)}}^{\mathrm{T}} \leqslant D^-$, 且当 $\det(D_{22})=0$ 时, 该不等式有严格小于号成立. 即, 在这种情况下, 估计方程估计比 Härdle 等 (1993) 给出的估计更有效.

因为 $\hat{\beta}$ 是 $\sqrt{n}$ 相合的, 所以可立刻得到下列结果.

定理 3.10.4　设条件 (C3.10.1)~(C3.10.5) 成立, 则

$$\sqrt{nh}\left[\hat{g}^*(u) - g_0(u) - \frac{1}{2}h^2\kappa_2 g''(u)\right] \xrightarrow{D} N\left(0, \nu_0\sigma^2 f^{-1}(u)\right),$$

其中 $\kappa_2 = \int u^2 K(u)\mathrm{d}u$, $\nu_0 = \int K^2(u)\mathrm{d}u$.

对误差方差的估计量 $\hat{\sigma}^2$, 我们有下述结果.

定理 3.10.5　设条件 (C3.10.1)~(C3.10.5) 成立, 且 $0<\mathrm{var}(\varepsilon^2)<\infty$, 则

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2)/[\mathrm{var}(\varepsilon^2)]^{1/2} \xrightarrow{D} N(0,1).$$

为应用定理 3.10.2 构造 β_0 的置信域, 需要 $\hat{\beta}$ 的一个 "嵌入" 估计. 因此需要估计 D. 由 "嵌入" 方法, 定义 D 的相合估计量为

$$\hat{D} = \frac{1}{n}\sum_{i=1}^{n}[X_i - \hat{\mu}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})][X_i - \hat{\mu}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})]^{\mathrm{T}}\hat{g}'(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})^2, \tag{3.10.9}$$

其中 $J_{\hat{\beta}^{(r)}}$ 是 $J_{\beta_0^{(r)}}$ 的估计量,

$$\hat{\mu}(u;\hat{\beta}) = \sum_{i=1}^{n} W_{ni}(u;\hat{\beta})X_i \tag{3.10.10}$$

是 $\mu(u)=E(X|\beta_0^{\mathrm{T}}X=u)$ 的估计量, $W_{ni}(u;\hat{\beta})$ 在 (3.10.6) 中定义.

由定理 3.10.2 和定理 3.10.5 可得

$$\left[\hat{\sigma}^2 J_{\hat{\beta}^{(r)}}(J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\hat{D}J_{\hat{\beta}^{(r)}})^- J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\right]^{-1/2}\sqrt{n}(\hat{\beta}-\beta_0)\xrightarrow{D}N(0,I_p).$$

再应用 Arnold (1981) 的定理 10.2d 即可得到

$$(\hat{\beta}-\beta_0)^{\mathrm{T}}\left[n^{-1}\hat{\sigma}^2 J_{\hat{\beta}^{(r)}}(J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\hat{D}J_{\hat{\beta}^{(r)}})^- J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\right]^{-1}(\hat{\beta}-\beta_0)\xrightarrow{D}\chi_p^2.$$

上述公式可以用来构造 β_0 的大样本置信域, 并可对 β_0 进行大样本假设检验.

需要指出, 如果使用修正的估计方程组

$$\begin{cases}\displaystyle\sum_{i=1}^n[Y_i-\hat{g}(\beta^{\mathrm{T}}X_i;\beta)]\hat{g}'(\beta^{\mathrm{T}}X_i;\beta)J_{\beta^{(r)}}^{\mathrm{T}}[X_i-\hat{\mu}(\beta^{\mathrm{T}}X_i)]=0,\\ ||\beta||-1=0,\end{cases}$$

那么可得到 β_0 的另一个估计量 $\hat{\beta}^*$. 该估计量也具有定理 3.10.2 给出的渐近正态性, 并且对带宽的条件可放宽到 $nh^4/\log n\to\infty$ 和 $nh^6\to 0$. 该条件包含了最优带宽 $O(n^{-1/5})$, 因此不需要对 $g(\cdot)$ 做欠光滑.

3.10.3 模拟研究

例3.10.1 考虑单指标模型

$$Y_i=20\exp\{-(\beta_1X_{i1}+\beta_2X_{i2})\}+\varepsilon_i,$$

其中 $(\beta_1,\beta_2)=(0.8,0.6)$, $\{(X_{i1},X_{i2}),1\leqslant i\leqslant n\}$ 是来自 $[-1,1]^2$ 上均匀分布的样本, $\{\varepsilon_i,1\leqslant i\leqslant n\}$ 是服从标准正态分布 $N(0,1)$ 的独立随机变量, $g(u)=20\exp(-u)$.

对 (β_1,β_2) 的估计, 我们使用两个方法: 3.10 节提出的估计方程 (EE) 方法和 3.2 节提出的最小二乘 (LS) 方法. 取权函数 $w(u)=I_{[0.01,1.39]}(u)$. 利用局部线性光滑方法估计 $g(\cdot)$, 核函数取为 $K(u)=0.75(1-u^2)_+$, 用交叉核实 (CV) 方法选择带宽 h. 分别在样本量 $n=30,60,100$ 下通过 500 次运行计算偏差 (Bias)、标准差 (SD) 和根均方误差 (RMSE). 模拟结果放在表 3.10.1 中.

从表 3.10.1 可以看出, β_1 和 β_2 的基于 EE 的估计比基于 LS 的估计有更小的根均方误差. 这就表明 EE 方法改进了估计结果. 此外, 偏差、标准差和根均方误差都随着样本量的增加而减少.

例3.10.2 为进一步说明我们的算法在高维情形的适应性, 考虑单指标模型

$$Y_i=60\exp\{-(\beta_0^{\mathrm{T}}X_i)^2\}+\varepsilon_i,$$

其中 $\beta_0=(1/3,2/3,2/3,0,\cdots,0)^{\mathrm{T}}$, $p=5,10,20$, X_i 由 $[0,1]^p$ 上的均匀分布产生, ε_i 由正态分布 $N(0,0.8^2)$ 产生, $g(u)=60\exp(-u^2)$. n 分别取为 60, 100 和 150.

表 3.10.1　例 3.10.1 的模拟结果. 估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的偏差、标准差和根均方误差

n	方法	$\hat{\beta}_1$			$\hat{\beta}_2$		
		Bias	SD	RMSE	Bias	SD	RMSE
30	EE	−0.00673	0.03805	0.03864	−0.00565	0.04968	0.05000
	LS	−0.00702	0.03837	0.03901	0.00608	0.04970	0.05007
60	EE	−0.00318	0.02777	0.02795	0.00243	0.03715	0.03723
	LS	−0.00315	0.02801	0.02819	0.00236	0.03757	0.03764
100	EE	−0.00153	0.02203	0.02208	0.00093	0.02918	0.02919
	LS	−0.00158	0.02239	0.02245	0.00100	0.02922	0.02924

取权函数 $w(u) = I_{[0.01,1.66]}(u)$, 用绝对内积 $|\hat{\beta}^{\mathrm{T}}\beta_0|$ 的均值 (Mean) 和标准差 (SD) 来度量估计量 $\hat{\beta}$ 的精度. 仍利用局部线性光滑方法估计 $g(\cdot)$, 取核函数 $K(u) = 0.75(1-u^2)_+$, 用交叉核实 (CV) 方法选择带宽 h. 表 3.10.2 给出的结果是基于 500 次重复运行. 从表 3.10.2 可以看出, EE 方法比 LS 方法有更大的内积值和更小的标准差. 两种方法的内积值都接近于 1, 其标准差都很小. 这说明 β_0 的估计量与其夹角的余弦值的精度很高, 但随着 p 的增加而降低.

表 3.10.2　例 3.10.2 的模拟结果. 估计值 $\hat{\beta}$ 和真实值 β_0 的绝对内积的均值和标准差

n	方法	$p=5$	$p=10$	$p=20$
		Mean (SD)	Mean (SD)	Mean (SD)
60	EE	0.99984 (0.00012)	0.99957 (0.00026)	0.99867 (0.00257)
	LS	0.99971 (0.00101)	0.99951 (0.00033)	0.99831 (0.00293)
100	EE	0.99991 (0.00007)	0.99977 (0.00017)	0.99947 (0.00029)
	LS	0.99987 (0.00039)	0.99968 (0.00142)	0.99925 (0.00208)
150	EE	0.99994 (0.00004)	0.99986 (0.00007)	0.99969 (0.00034)
	LS	0.99990 (0.00066)	0.99982 (0.00051)	0.99950 (0.00218)

3.10.4　定理的证明

在本节中, 使用 c 表示任何正的常数, 每次出现时可以代表不同的值. 为证明上节的定理, 首先给出几个引理. 记 $\mathcal{A}_n = \{(u,\beta) : (u,\beta) \in \mathcal{U} \times \mathcal{B}_n^*\}$.

引理 3.10.1　设 $\xi_1(u,\beta), \cdots, \xi_n(u,\beta)$ 为随机变量序列. 记 $f_{u,\beta}(V_i) = \xi_i(u,\beta)$, $i = 1, \cdots, n$, 其中 $V_1, \cdots, V_n$ 为随机变量序列, $f_{u,\beta}(\cdot)$ 是 $\mathcal{A}_n$ 上的函数. 又设对常数 x^*, β^*, $a > 0$ 和 $c > 0$, $f_{u,\beta}(\cdot)$ 满足

$$\frac{1}{n}\sum_{i=1}^{n}\left|f_{u,\beta}(V_i) - f_{u^*,\beta^*}(V_i)\right| \leqslant cn^a\left[||\beta-\beta^*|| + |u-u^*|\right]. \tag{3.10.11}$$

如果对 $(u,\beta)\in\mathcal{A}_n$ 和仅依赖于 n 的 $\varepsilon_n>0$,

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(u,\beta)\right|>\frac{1}{2}\varepsilon_n\right\}\leqslant\frac{1}{2}, \tag{3.10.12}$$

则有常数 $c_0>0$, 使得

$$\begin{aligned}&P\left\{\sup_{(u,\beta)\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(u,\beta)\right|>\frac{1}{2}\varepsilon_n\right\}\\&\leqslant c_0n^{2pa}\varepsilon_n^{-2p}E\left[\sup_{(u,\beta)\in\mathcal{A}_n}2\exp\left(\frac{-n^2\varepsilon_n^2/128}{\sum\limits_{i=1}^{n}\xi_i^2(u,\beta)}\right)\wedge 1\right].\end{aligned} \tag{3.10.13}$$

证 设 $\{\xi_1^*(u,\beta),\cdots,\xi_n^*(u,\beta)\}$ 是 $\{\xi_1(u,\beta),\cdots,\xi_n(u,\beta)\}$ 的独立翻版. 现产生独立符号随机变量 $\sigma_1,\cdots,\sigma_n$, 这里 $P\{\sigma_i=1\}=P\{\sigma_i=-1\}=\dfrac{1}{2}$, $\{\sigma_i,1\leqslant i\leqslant n\}$ 与 $\{\xi_i(u,\beta),\xi_i^*(u,\beta),1\leqslant i\leqslant n\}$ 独立. 由对称性, $\sigma_i(\xi_i-\xi_i^*)$ 与 $(\xi_i-\xi_i^*)$ 同分布. 由 Pollard(1984) 的对称性引理可推出

$$\begin{aligned}&P\left\{\sup_{(u,\beta)\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(u,\beta)\right|>\varepsilon_n\right\}\\&\leqslant 2P\left\{\sup_{(u,\beta)\in_n}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\xi_i(u,\beta)-\xi_i^*(u,\beta)\right]\right|>\frac{1}{2}\varepsilon_n\right\}\\&=2P\left\{\sup_{(u,\beta)\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\left[\xi_i(u,\beta)-\xi_i^*(u,\beta)\right]\right|>\frac{1}{2}\varepsilon_n\right\}\\&\leqslant 4P\left\{\sup_{(u,\beta)\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\xi_i(u,\beta)\right|>\frac{1}{4}\varepsilon_n\right\}.\end{aligned} \tag{3.10.14}$$

用 P_n 表示 $V_1,\cdots,V_n$ 的经验测度. 设 $\mathcal{F}=\{f_{u,\beta}(\cdot):|u|\leqslant c,\|\beta\|\leqslant B\}$ 是由 u 和 β 构成的函数 $f_{u,\beta}(V_i)=\xi_i(u,\beta)$ 的函数类. 记 $\mathcal{V}=(V_1,\cdots,V_n)$. 给定 $\mathcal{V}$, 在 $\mathcal{F}$ 中选择函数 $f_1^\circ,\cdots,f_m^\circ$, 使得对 $\mathcal{F}$ 中的每个 $f_{u,\beta}$,

$$\min_{j\in\{1,\cdots,m\}}\frac{1}{n}\sum_{i=1}^{n}|f_{u,\beta}(V_i)-f_j^\circ(V_i)|<\varepsilon_n. \tag{3.10.15}$$

设 $N(\varepsilon_n,P_n,\mathcal{F})$ 为满足 (3.10.15) 的所有集的最小的 m. 用 $f_{u,\beta}^*$ 表示使 f_j° 达到最

小值的函数. 那么

$$\begin{aligned}
&P\left\{\sup_{(u,\beta)\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{i=1}^n\sigma_i\xi_i(u,\beta)\right|>\frac{1}{4}\varepsilon_n\middle|\mathcal{V}\right\}\\
\leqslant &P\left\{\sup_{(u,\beta)\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{i=1}^n\sigma_i f_{u,\beta}(V_i)\right|>\frac{1}{4}\varepsilon_n\middle|\mathcal{V}\right\}\\
\leqslant &P\left\{\sup_{(u,\beta)\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{i=1}^n\sigma_i f^*_{u,\beta}(V_i)\right|>\frac{1}{8}\varepsilon_n\middle|\mathcal{V}\right\}\\
\leqslant &N(\varepsilon_n,P_n,\mathcal{F})\max_{j\in\{1,\cdots,N\}}P\left\{\left|\frac{1}{n}\sum_{i=1}^n\sigma_i f^\circ_j(V_i)\right|>\frac{1}{8}\varepsilon_n\middle|\mathcal{V}\right\}. \qquad (3.10.16)
\end{aligned}$$

我们现在需要确定 $N(\varepsilon_n,P_n,\mathcal{F})$ 的阶. 对每个满足 (3.10.15) 式的集, 每个 f°_j 都有一个数对 (u_j,β_j), 使得 $f^\circ_j(v)=f_{u_j,\beta_j}(v)$. 那么对所有 $(u,\beta)\in\mathcal{A}_n$, 由 (3.10.11) 式可得

$$\frac{1}{n}\sum_{i=1}^n\left|f_{u,\beta}(V_i)-f_{u_j,\beta_j}(V_i)\right|\leqslant cn^a(\|\beta-\beta_j\|+|u-u_j|).$$

我们希望用 ε_n 界定上式右端. 为此, 对每个 $(u,\beta)\in\mathcal{A}_n$, 在以 (u,β) 为中心半径为 $r_n=O(n^{-a}\varepsilon_n)$ 的球内需要一个数对 (u_j,β_j). 因此, 满足 (3.10.15) 式的数 N 可用 $r_n^{-p}r_n^{-p}=cn^{2pa}\varepsilon_n^{-2p}$ 界定, 即

$$N(\varepsilon_n,P_n,\mathcal{F})\leqslant cn^{2pa}\varepsilon_n^{-2p}. \qquad (3.10.17)$$

在给定 $\mathcal{V}$ 的条件下, $\sigma_i f^\circ_j(V_i)$ 有界. 因此, 利用 Hoeffding 不等式 (Hoeffding, 1963) 可得

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n\sigma_i f^\circ_j(V_i)\right|>\frac{1}{8}\varepsilon_n\middle|\mathcal{V}\right\}\leqslant 2\exp\left(\frac{-2n(\varepsilon_n/8)^2}{\sum_{i=1}^n 4f^2_{u_j,\beta_j}(V_i)}\right)\wedge 1. \qquad (3.10.18)$$

结合 (3.10.14)、(3.10.16)、(3.10.17) 和 (3.10.18) 式即证得 (3.10.13) 式.

引理 3.10.2　设条件 (C3.10.1)~(C3.10.4) 成立. 如果对任何 $0<a<1/2$ 和 $c>0$, $h=cn^{-a}$, 则对任何 $1\leqslant i\leqslant n$ 和接近 β_0 的点 β(包括 β_0), 有

$$E\left[\sum_{j=1}^n W_{nj}(\beta^{\mathrm{T}}X_i;\beta)g(\beta^{\mathrm{T}}X_j)-g(\beta^{\mathrm{T}}X_i)\right]^2=O(h^4),$$

$$E\left[\sum_{j=1}^n \widetilde{W}_{nj}(\beta^{\mathrm{T}}X_i;\beta)g(\beta^{\mathrm{T}}X_j)-g'(\beta^{\mathrm{T}}X_i)\right]^2=O(h_1^2),$$

$$E\left[\sum_{j=1}^{n} W_{ni}(\beta^{\mathrm{T}} X_j;\beta)\varphi(\beta^{\mathrm{T}} X_j) - \varphi(\beta^{\mathrm{T}} X_i)\right]^2 = O(\sqrt{h}\ ),$$

其中 $\varphi(u) = g'(u)\mu_s(u)$, $\mu_s(u)$ 为 $\mu(u) = E(X|X^{\mathrm{T}}\beta = u)$ 的第 s 个分量.

证 仅证第一个等式, 类似可证明其他两个等式. 记 $U_i = \beta^{\mathrm{T}} X_i$, 用 $S_{n,l}(U_i)$ 和 $V_{nj}(U_i)$ 分别代替 $S_{n,l}(\beta^{\mathrm{T}} X_i;\beta,h)$ 和 $V_{nj}(\beta^{\mathrm{T}} X_i;\beta,h)$. 简单计算可得

$$g(U_i) - \sum_{j=1}^{n} W_{nj}(U_i)g(U_j) = \frac{\sum_{j=1}^{n} V_{nj}(U_i)[g(U_j) - g(U_i)]}{\sum_{j=1}^{n} V_{nj}(U_i)}. \tag{3.10.19}$$

用 $E_{U_i}[\cdot]$ 表示给定 U_i 下的条件期望. 若 $E|Z_n|^r = O(a_n^r)$, 则记 $Z_n = O_r(a_n)$. 利用 Cauchy-Schwarz 不等式容易得到

$$O_r(a_n)O_r(b_n) = O_{r/2}(a_n b_n), \tag{3.10.20}$$

$$Z_n = E_{U_i}[Z_n] + O_r\big((E|Z_n - E_{U_i}[Z_n]|^4)^{1/r}\big). \tag{3.10.21}$$

由条件 (C3.10.1) 可知 $M_f = \sup_t f(u) < \infty$, 且存在常数 $L > 0$ 使得对任何 y 和 u, $|f(y) - f(u)| \leqslant L|y - u|$. 对 $i = 1, \cdots, n$, 当 n 充分大可得

$$E_{U_i}\big[S_{n,l}(U_i)\big] = h^l \kappa_l f(U_i)[1 + O(h)], \quad l = 0,1,2, \tag{3.10.22}$$

其中 $\kappa_l = \int_{-\infty}^{\infty} u^l K(u)\mathrm{d}u$, $l = 0,1,2$, $\kappa_0 = 1$, $\kappa_1 = 0$.

由独立随机变量和的不等式 (Petrov, 1991), 可得

$$\begin{aligned}
&E\{|S_{n,l}(U_i) - E_{U_i}[S_{n,l}(U_i)]|^4\} \\
&= E\{E_{U_i}\big[|S_{n,l}(U_i) - E_{U_i}S_{n,l}(U_i)|^4\big]\} \\
&\leqslant cn^{-4}(n-1)h^{4l-3}\int_{-\infty}^{\infty}|u^l K(u)|^4\mathrm{d}u[Ef(U_i) + Lh] \\
&\quad + cn^{-4}E\left\{2(n-1)h^{2l-3}\int_{-\infty}^{\infty}u^{2l}K^2(u)\mathrm{d}u[f(U_i) + Lh]\right\}^2 \\
&= O(n^{-3}h^{4l-3}) + O(n^{-2}h^{2l-2}).
\end{aligned}$$

结合 (3.10.21) 和 (3.10.22) 式可证得

$$
\begin{aligned}
S_{n,l}(U_i) &= E_{U_i}[S_{n,l}(U_i)] + O_4(h^l(nh)^{-1/2}) \\
&= \kappa_l h^l f(U_i)(1 + O_4(h + (nh)^{-1/2})), \quad l = 0, 1, 2.
\end{aligned}
\tag{3.10.23}
$$

由 (3.10.23) 和 (3.10.20) 式可得

$$
\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n} V_{nj}(U_i) &= S_{n,0}(U_i)S_{n,2}(U_i) - S_{n,1}^2(U_i) \\
&= h^2\kappa_2 f^2(U_i)(1 + O_2(h + (nh)^{-1/2})).
\end{aligned}
\tag{3.10.24}
$$

记 $V_n(u) = (nh^2)^{-1}\sum_{j=1}^{n} V_{nj}(u)$, $V(u) = \kappa_2 f^2(u)$. 利用引理 3.10.1、(3.10.24) 式和 Borel-Cantelli 引理, 不难证得

$$
\sup_{u\in\mathcal{U}} |V_n(u) - V(u)| \to 0, \quad \text{a.s.} \tag{3.10.25}
$$

由条件 (C3.10.1) 可知 $\inf_{u\in\mathcal{U}} f(u) \geqslant c_1 > 0$. 因此, 当 n 充分大时,

$$
\inf_{u\in\mathcal{U}} |V_n(u)| \geqslant \inf_{u\in\mathcal{U}} |V(u)| - \sup_{u\in\mathcal{U}} |V_n(u) - V(u)| > \kappa_2 c_1^2/2 > 0, \quad \text{a.s.} \tag{3.10.26}
$$

记 $H(U_j, U_i) = g(U_j) - g(U_i) + g'(U_i)(U_j - U_i)$. 注意到 $\sum_{j=1}^{n} V_{nj}(U_i)(U_j - U_i) = 0$, 有

$$
\begin{aligned}
&\sum_{j=1}^{n} V_{nj}(U_i)[g(U_j) - g(U_i)] = \sum_{j=1}^{n} H(U_j, U_i)V_{nj}(U_i) \\
&= \sum_{j=1}^{n} H(U_j, U_i)K_h(U_j - U_i)S_{n,2}(U_i) \\
&\quad - \sum_{j=1}^{n} H(U_j, U_i)(U_j - U_i)K_h(U_j - U_i)S_{n,1}(U_i).
\end{aligned}
\tag{3.10.27}
$$

类似于 (3.10.23) 式的证明, 对 $l = 0, 1$, 有

$$
\begin{aligned}
&\frac{1}{nh^{2+l}}\sum_{j=1}^{n} H(U_j, U_i)(U_j - U_i)^l K_h(U_j - U_i) \\
&= h^{-2-l}E_{U_i}[H(U_1, U_i)(U_1 - U_i)^l K_h(U_1 - U_i)] + o_4(1) \\
&\equiv d_{nl} + o_4(1), \quad 2 \leqslant i \leqslant n
\end{aligned}
\tag{3.10.28}
$$

$$|d_{nl}| \leqslant c\int_{-\infty}^{\infty}|u|^{l+2}K(u)\mathrm{d}u[f(U_i)+O(h)]=O(1).$$

结合 (3.10.23)、(3.10.27) 和 (3.10.28) 式可证得

$$\sum_{j=1}^{n}V_{nj}(U_i)[g(U_j)-g(U_i)]=nh^4f(U_i)\kappa_2d_{n0}+o_2(nh^4).$$

再结合 (3.10.26) 式可得

$$E\left|\sum_{j=1}^{n}V_{nj}(U_i)[g(U_j)-g(U_i)]\bigg/\sum_{j=1}^{n}V_{nj}(U_i)\right|^2$$

$$\leqslant c(nh^2)^{-2}E\left|\sum_{j=1}^{n}V_{nj}(U_i)[g(U_j)-g(U_i)]\right|^2=O(h^4). \qquad (3.10.29)$$

故由 (3.10.19) 和 (3.10.29) 式即证明了引理 3.10.2 的第一个等式.

引理 3.10.3 设条件 (C3.10.1)~(C3.10.4) 成立. 如果对任何 $0<a<1/2$ 和 $c>0$, $h=cn^{-a}$, 则对任何 $1\leqslant i\leqslant n$ 和接近 β_0 的点 β(包括 β_0), 有

$$\begin{cases} E\{W_{ni}^2(\beta^{\mathrm{T}}X_i;\beta)\}=O((nh)^{-2}), \\ E\left[\sum\limits_{j=1,j\neq i}^{n}W_{nj}^2(\beta^{\mathrm{T}}X_i;\beta)\right]=O((nh)^{-1}), \end{cases}$$

$$\begin{cases} E\{\widetilde{W}_{ni}^2(\beta^{\mathrm{T}}X_i;\beta)\}=O((nh_1)^{-2}+(n^3h_1^5)^{-1}), \\ E\left[\sum\limits_{j=1,j\neq i}^{n}\widetilde{W}_{nj}^2(\beta^{\mathrm{T}}X_i;\beta)\right]=O((nh_1^3)^{-1}). \end{cases}$$

证 类似于引理 3.10.2 的证明即证, 这里证明省略.

定理 3.10.1 的证 仅证第一式, 类似的证明可证第二式. 记 $\tilde{g}(X_i,\varepsilon_i)=g(\beta_0^{\mathrm{T}}X_i)-g(u)+\varepsilon_i$, $i=1,\cdots,n$. 简单计算可得

$$\hat{g}(u;\beta)-g(u)=\sum_{i=1}^{n}W_{ni}(u;\beta)\tilde{g}(X_i,\varepsilon_i).$$

记 $\xi_i(u,\beta)=n(nh/\log n)^{1/2}W_{ni}(u;\beta)\tilde{g}(X_i,\varepsilon_i)$, $f_{u,\beta}(V_i)=\xi_i(u,\beta)$, $V_i=(X_i,\varepsilon_i)$, $i=1,\cdots,n$. 这里使用引理 3.10.1 来证明该定理, 那么需要验证条件 (3.10.11) 和 (3.10.12) 式. 简单计算可知 (3.10.11) 式成立, 下面验证条件 (3.10.12) 式. 由引理 3.10.2 和引理 3.10.3 可得

$$E[\hat{g}(u;\beta)-g(u)]^2=E\left[\sum_{i=1}^{n}W_{ni}(u;\beta)\tilde{g}(X_i,\varepsilon_i)\right]^2$$
$$\leqslant cE\left[\sum_{i=1}^{n}W_{ni}(u;\beta)g(\beta^{\mathrm{T}}X_i)-g(u)\right]^2+cE\left[\sum_{i=1}^{n}W_{ni}^2(u;\beta)\right]$$
$$\leqslant ch^4+c(nh)^{-1}. \tag{3.10.30}$$

给定 $M>0$, 由 Chevbychev 不等式和 (3.10.30) 式, 可得

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(u,\beta)\right|>\frac{1}{2}M\right\}\leqslant 4M^{-2}E\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i(u,\beta)\right]^2$$
$$\leqslant 4M^{-2}nh(\log n)^{-1}E\left[\sum_{i=1}^{n}W_{ni}(u;\beta)\tilde{g}(X_i,\varepsilon_i)\right]^2$$
$$\leqslant cM^{-2}[cnh^5+c(\log n)^{-1}]. \tag{3.10.31}$$

因此, 由条件 (C3.10.5), 可选择 M 充分大以致 (3.10.31) 式右边 $\leqslant\frac{1}{2}$. 因此, (3.10.12) 式成立. 由引理 3.10.3 可得

$$n^{-2}\sum_{i=1}^{n}E[\xi_i^2(u,\beta)]=nh(\log n)^{-1}\sum_{i=1}^{n}E\left[W_{ni}(u;\beta)\tilde{g}(X_i,\varepsilon_i)\right]^2$$
$$\leqslant cnh(\log n)^{-1}\sum_{i=1}^{n}E\left[W_{ni}^2(u;\beta)\right]\leqslant c(\log n)^{-1}.$$

这就推出

$$n^{-2}\sum_{i=1}^{n}\xi_i^2(u,\beta)=O_P((\log n)^{-1}).$$

因此, 由引理 3.10.1 可得

$$P\left\{\sup_{(u,\beta)\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(u,\beta)\right|>\frac{1}{2}M\right\}\leqslant cn^{2pa}M^{-2p}\exp(-cM^2\log n). \tag{3.10.32}$$

当 M 充分大时, (3.10.32) 式右边 $\leqslant n^{-2}$. 故由 Borel-Cantelli 引理即证得定理 3.10.1 的第一式.

引理 3.10.4 设条件 (C3.10.1)~(C3.10.5) 成立, 则

$$\sup_{\beta^{(r)}\in\mathcal{B}_n}||R(\beta^{(r)})-U(\beta_0^{(r)})+nJ_{\beta_0^{(r)}}^{\mathrm{T}}DJ_{\beta_0^{(r)}}(\beta^{(r)}-\beta_0^{(r)})||=o_P(\sqrt{n}),$$

其中 $\mathcal{B}_n = \{\beta^{(r)} : ||\beta^{(r)} - \beta_0^{(r)}|| \leqslant cn^{-1/2}\}$, $c > 0$ 为常数, D 在定理 3.10.1 中定义,

$$R(\beta^{(r)}) = \sum_{i=1}^{n}[Y_i - \hat{g}(\beta^{\mathrm{T}} X_i; \beta)]\hat{g}'(\beta^{\mathrm{T}} X_i; \beta) J_{\beta^{(r)}}^{\mathrm{T}} X_i,$$

$$U(\beta_0^{(r)}) = \sum_{i=1}^{n} \varepsilon_i g'(\beta_0^{\mathrm{T}} X_i) J_{\beta_0^{(r)}}^{\mathrm{T}} [X_i - E(X_i|\beta_0^{\mathrm{T}} X_i)].$$

证 将 $R(\beta^{(r)})$ 分解, 可得

$$\begin{aligned} R(\beta^{(r)}) =& \sum_{i=1}^{n} \varepsilon_i g'(\beta_0^{\mathrm{T}} X_i) J_{\beta^{(r)}}^{\mathrm{T}} [X_i - E(X_i|\beta_0^{\mathrm{T}} X_i)] \\ &+ \sum_{i=1}^{n} \varepsilon_i [\hat{g}'(\beta^{\mathrm{T}} X_i; \beta) - g'(\beta_0^{\mathrm{T}} X_i)] J_{\beta^{(r)}}^{\mathrm{T}} X_i \\ &- \sum_{i=1}^{n} g'(\beta_0^{\mathrm{T}} X_i) J_{\beta^{(r)}}^{\mathrm{T}} X_i [\hat{g}(\beta^{\mathrm{T}} X_i; \beta) - \hat{g}(\beta_0^{\mathrm{T}} X_i; \beta_0)] \\ &- \sum_{i=1}^{n} g'(\beta_0^{\mathrm{T}} X_i) J_{\beta^{(r)}}^{\mathrm{T}} \{X_i[\hat{g}(\beta_0^{\mathrm{T}} X_i; \beta_0) - g(\beta_0^{\mathrm{T}} X_i)] - \varepsilon_i \mu(\beta_0^{\mathrm{T}} X_i)\} \\ &- \sum_{i=1}^{n} [\hat{g}(\beta^{\mathrm{T}} X_i; \beta) - g(\beta_0^{\mathrm{T}} X_i)][\hat{g}'(\beta^{\mathrm{T}} X_i; \beta) - g'(\beta_0^{\mathrm{T}} X_i)] J_{\beta^{(r)}}^{\mathrm{T}} X_i \\ \equiv& R_1(\beta^{(r)}) + R_2(\beta^{(r)}) - R_3(\beta^{(r)}) - R_4(\beta^{(r)}) - R_5(\beta^{(r)}). \end{aligned} \tag{3.10.33}$$

注意到对所有 $\beta^{(r)} \in \mathcal{B}_n$, $J_{\beta^{(r)}} - J_{\beta_0^{(r)}} = O_P(n^{-1/2})$, 因此, 由大数定律可得

$$\sup_{\beta^{(r)} \in \mathcal{B}_n} ||R_1(\beta^{(r)}) - U(\beta_0^{(r)})|| = o_P(\sqrt{n}). \tag{3.10.34}$$

下面证明

$$\sup_{\beta^{(r)} \in \mathcal{B}_n} ||R_2(\beta^{(r)})|| = o_P(\sqrt{n}). \tag{3.10.35}$$

需要应用引理 3.10.1. 这里去掉 u 后仍可以使用引理 3.10.1. 记

$$\xi_i(\beta) = \sqrt{n}\varepsilon_i[\hat{g}'(\beta^{\mathrm{T}} X_i, \beta) - g'(\beta_0^{\mathrm{T}} X_i)]X_{is},$$
$$f_\beta(V_i) = \xi_i(\beta), \qquad V_i = (X_i, \varepsilon_i), \quad i = 1, \cdots, n.$$

其中 X_{is} 为 X_i 的第 s 个分量. 下面验证引理 3.10.1 的条件 (3.10.11) 和 (3.10.12). 由关于核函数的条件 (C3.10.4) 可得

$$\frac{1}{n}\sum_{i=1}^{n} |f_\beta(V_i) - f_{\beta^*}(V_i)| \leqslant cn^{5/2}h^{-2}||\beta - \beta^*|| = cn^a||\beta - \beta^*||,$$

其中 $a=\frac{5}{2}+2\lambda\left(\frac{1}{5}\leqslant\lambda<\frac{1}{2}\right)$. 由于对所有 $\beta^{(r)}\in\mathcal{B}_n$, $||\beta^{(r)}-\beta_0^{(r)}||\leqslant cn^{-1/2}$ 蕴涵 $||\beta-\beta_0||\leqslant cn^{-1/2}$. 因此, (3.10.11) 式成立.

下面验证 (3.10.12) 式. 由条件 (C3.10.5)、引理 3.10.2 和引理 3.10.3 可得

$$
\begin{aligned}
&E\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i(\beta)\right]^2\\
&\leqslant 2n^{-1}\sum_{i=1}^{n}E\left[\sum_{j=1}^{n}\widetilde{W}_{nj}(\beta^{\mathrm{T}}X_i;\beta)g(\beta_0^{\mathrm{T}}X_j)-g'(\beta_0^{\mathrm{T}}X_i)\right]^2\\
&\quad+2n^{-1}\sum_i\sum_j\sum_k\sum_l E\left[\widetilde{W}_{nj}(\beta^{\mathrm{T}}X_i;\beta)\widetilde{W}_{nl}(X_k^{\mathrm{T}}\beta;\beta)\varepsilon_i\varepsilon_k\varepsilon_j\varepsilon_l\right]\\
&\leqslant ch_1^2+cn^{-1}\left\{\sum_{i=1}^{n}E[\widetilde{W}_{ni}^2(\beta^{\mathrm{T}}X_i;\beta)]+\sum_{i\neq j}E[\widetilde{W}_{nj}^2(\beta^{\mathrm{T}}X_i;\beta)]\right\}\\
&\leqslant ch_1^2+c(nh_1^3)^{-1}\to 0.
\end{aligned}
$$

因此, 可以推出: 当 n 充分大时,

$$
P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(\beta)\right|>\frac{1}{2}\varepsilon\right\}\leqslant ch_1^2+c(nh_1^3)^{-1}<\frac{1}{2}.
$$

这就验证了 (3.10.12) 式成立. 由定理 3.10.1 可得

$$
\frac{1}{n^2}\sum_{i=1}^{n}\xi_i^2(\beta)=\sup_{u\in\mathcal{U},\beta\in\mathcal{B}_n^*}[\hat{g}'(x^{\mathrm{T}}\beta;\beta)-g'(x^{\mathrm{T}}\beta_0)]^2O_P(1)=O_P\left((nh_1^3/\log n)^{-1}\right).
$$

由引理 3.10.1, 当 $nh_1^3/\log^2 n\to\infty$, 有

$$
P\left\{\sup_{(u,\beta)\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(\beta)\right|>\frac{1}{2}\varepsilon\right\}\leqslant cn^{2pa}\varepsilon^{-2p}\exp\left(-\frac{cnh_1^3}{\log n}\right)\to 0.
$$

这就证明了 (3.10.35) 式.

下面考虑 $R_3(\beta^{(r)})$. 定义 $g(\beta^{\mathrm{T}}X;\beta)=E(Y|\beta^{\mathrm{T}}X)$. 显然 $g(\beta_0^{\mathrm{T}}X;\beta_0)=g(\beta_0^{\mathrm{T}}X)$. 经过计算可得

$$
R_3(\beta^{(r)})=\sum_{i=1}^{n}g'(\beta_0^{\mathrm{T}}X_i)J_{\beta^{(r)}}^{\mathrm{T}}X_i[g(\beta^{\mathrm{T}}X_i;\beta)-g(\beta_0^{\mathrm{T}}X_i)]
$$

$$+\sum_{j=1}^{n}\varepsilon_j\left[\sum_{i=1}^{n}W_{nj}(\beta^{\mathrm{T}}X_i;\beta)g'(\beta_0^{\mathrm{T}}X_i)J_{\beta^{(r)}}^{\mathrm{T}}X_i-g'(\beta^{\mathrm{T}}X_j)J_{\beta^{(r)}}^{\mathrm{T}}\mu(\beta^{\mathrm{T}}X_j)\right]$$

$$-\sum_{j=1}^{n}\varepsilon_j\left[\sum_{i=1}^{n}W_{nj}(\beta_0^{\mathrm{T}}X_i;\beta_0)g'(\beta_0^{\mathrm{T}}X_i)J_{\beta^{(r)}}^{\mathrm{T}}X_i-g'(\beta_0^{\mathrm{T}}X_j)J_{\beta^{(r)}}^{\mathrm{T}}\mu(\beta_0^{\mathrm{T}}X_j)\right]$$

$$+\sum_{i=1}^{n}g'(\beta_0^{\mathrm{T}}X_i)J_{\beta^{(r)}}^{\mathrm{T}}X_i\sum_{j=1}^{n}W_{nj}(\beta^{\mathrm{T}}X_i;\beta)[g(\beta^{\mathrm{T}}X_j;\beta)-g(\beta^{\mathrm{T}}X_i;\beta)]$$

$$+\sum_{j=1}^{n}\varepsilon_j\left[g'(\beta^{\mathrm{T}}X_j)J_{\beta^{(r)}}^{\mathrm{T}}\mu(\beta^{\mathrm{T}}X_j)-g'(\beta_0^{\mathrm{T}}X_j)J_{\beta^{(r)}}^{\mathrm{T}}\mu(\beta_0^{\mathrm{T}}X_j)\right]$$

$$+\sum_{i=1}^{n}g'(\beta_0^{\mathrm{T}}X_i)J_{\beta^{(r)}}^{\mathrm{T}}X_i\sum_{j=1}^{n}W_{nj}(\beta_0^{\mathrm{T}}X_i;\beta_0)[g(\beta_0^{\mathrm{T}}X_j)-g(\beta_0^{\mathrm{T}}X_i)]$$

$$+\sum_{i=1}^{n}g'(\beta_0^{\mathrm{T}}X_i)J_{\beta^{(r)}}^{\mathrm{T}}X_i\sum_{j=1}^{n}W_{nj}(\beta^{\mathrm{T}}X_i;\beta)[g(\beta^{\mathrm{T}}X_j;\beta)-g(\beta_0^{\mathrm{T}}X_j)]$$

$$\equiv\sum_{\nu=1}^{7}R_{3\nu}(\beta^{(r)}).$$

由 Taylor 公式和大数定律容易得到

$$\sup_{\beta^{(r)}\in\mathcal{B}_n}||R_{31}(\beta^{(r)})-nJ_{\beta_0^{(r)}}^{\mathrm{T}}DJ_{\beta_0^{(r)}}(\beta^{(r)}-\beta_0^{(r)})||=o_P(\sqrt{n}).$$

利用引理 3.10.1~ 引理 3.10.3, 可以证明

$$\sup_{\beta^{(r)}\in\mathcal{B}_n}||R_{3\nu}(\beta^{(r)})||=o_P(\sqrt{n}),\quad \nu=2,\cdots,7.$$

结合上述证明可得

$$\sup_{\beta^{(r)}\in\mathcal{B}_n}||R_3(\beta^{(r)})-nJ_{\beta_0^{(r)}}^{\mathrm{T}}DJ_{\beta_0^{(r)}}(\beta^{(r)}-\beta_0^{(r)})||=o_P(\sqrt{n}).\tag{3.10.36}$$

现在考虑 $R_4(\beta^{(r)})$. 记 $R_4(\beta^{(r)})=J_{\beta^{(r)}}^{\mathrm{T}}R_4^*(\beta^{(r)})$. 用 $R_{4,s}^*$ 表示 $R_4^*(\beta^{(r)})$ 的第 s 个分量. 由引理 3.10.2 和引理 3.10.3 可得

$$\begin{aligned}n^{-1}E(R_{4,s}^{*2})&\leqslant cn^{-1}\sum_{i=1}^{n}E\left[\sum_{j=1}^{n}W_{ni}(\beta_0^{\mathrm{T}}X_j;\beta_0)g'(\beta_0^{\mathrm{T}}X_j)X_{js}-g'(\beta_0^{\mathrm{T}}X_i)\mu_s(\beta_0^{\mathrm{T}}X_i)\right]^2\\&\quad+c\sum_{i=1}^{n}E\left[\sum_{j=1}^{n}W_{nj}(\beta_0^{\mathrm{T}}X_i;\beta_0)g(\beta_0^{\mathrm{T}}X_j)-g(\beta_0^{\mathrm{T}}X_i)\right]^2\\&\leqslant c(nh)^{-1}+c\sqrt{h}+cnh^4\to 0.\end{aligned}$$

这就推出

$$\sup_{\beta^{(r)}\in\mathcal{B}_n}||R_4(\beta^{(r)})|| = o_P(\sqrt{n}). \tag{3.10.37}$$

利用定理 3.10.1 不难得到

$$\sup_{\beta^{(r)}\in\mathcal{B}_n}||R_5(\beta^{(r)})|| = o_P(\sqrt{n}). \tag{3.10.38}$$

将 (3.10.34)~(3.10.38) 式代入 (3.10.33) 式, 即可证得引理 3.10.4.

定理 3.10.2 的证 该定理的证明分为两步: 步骤 (I) 证明估计方程组 (3.10.7) 的解 $\hat{\beta}$ 的存在性; 步骤 (II) 证明 $\hat{\beta}$ 的渐近正态性.

(I) *存在性的证*. 首先证明下列事实: 在条件 (C3.10.1)~(C3.10.5) 下, 估计方程组 (3.10.7) 以概率 1 存在一个解 $\hat{\beta}\in\mathcal{B}_{1n}$, 其中 $\mathcal{B}_{1n}=\{\beta:||\beta-\beta_0||=B_1n^{-1/2}\}$, $0<B_1<\infty$ 为某个常数.

事实上, 由 (3.10.2) 式, 仅需证明在 $\mathcal{B}_{2n}$ 中估计方程 $R(\beta^{(r)})=0$ 的解的存在性, 其中 $\mathcal{B}_{2n}=\{\beta^{(r)}:||\beta^{(r)}-\beta_0^{(r)}||=B_2n^{-1/2}\}$, $0<B_2<\infty$ 为某个常数. 对确定的常数 B_2 及任意 $\beta^{(r)}\in\mathcal{B}_{2n}$, 由引理 3.10.4 可得

$$\begin{aligned}(\beta^{(r)}-\beta_0^{(r)})^{\mathrm{T}}R(\beta^{(r)}) = &(\beta^{(r)}-\beta_0^{(r)})^{\mathrm{T}}U(\beta_0^{(r)})\\ &-n(\beta^{(r)}-\beta_0^{(r)})^{\mathrm{T}}A(\beta^{(r)}-\beta_0^{(r)})+o_P(1),\end{aligned} \tag{3.10.39}$$

其中 $A=J_{\beta_0^{(r)}}^{\mathrm{T}}DJ_{\beta_0^{(r)}}$. 下列论证类似于 Weisberg 和 Welsh(1994) 使用的方法, 其中使用了 Ortega 和 Rheinboldt(1973) 的 (6.3.4) 式. 注意到 (3.10.39) 式受 B_2^2 控制, 因为 $\sqrt{n}||\beta^{(r)}-\beta_0^{(r)}||=B_2$, 而 $|(\beta^{(r)}-\beta_0^{(r)})^{\mathrm{T}}U(\beta_0^{(r)})|=B_2O_P(1)$ 且 $n(\beta^{(r)}-\beta_0^{(r)})^{\mathrm{T}}A(\beta^{(r)}-\beta_0^{(r)})\simeq B_2^2$. 因此, 对任意给定的 $\eta>0$, 如果选择 B_2 足够大, 那么以概率 $1-\eta$, 有 $(\beta^{(r)}-\beta_0^{(r)})^{\mathrm{T}}R(\beta_0^{(r)})<0$. 然后, 类似于 Welsh(1989) 的定理 5.1 的证明即可证得的估计方程 $R(\beta^{(r)})=0$ 在 $\mathcal{B}_{2n}$ 中的解存在. 这里省略其证明细节.

(II) *渐近正态性的证*. 由 (I) 可知 $\hat{\beta}^{(r)}$ 是方程 $R(\beta^{(r)})=0$ 在 $\mathcal{B}_{2n}$ 中的一个解, 即, $R(\hat{\beta}^{(r)})=0$. 因此, 由引理 3.10.4 可得

$$0=nU(\beta_0^{(r)})-J_{\beta_0^{(r)}}^{\mathrm{T}}DJ_{\beta_0^{(r)}}(\hat{\beta}^{(r)}-\beta_0^{(r)})+o_P(n^{-1/2}).$$

由上式可得

$$\sqrt{n}(\hat{\beta}^{(r)}-\beta_0^{(r)})=(J_{\beta_0^{(r)}}^{\mathrm{T}}DJ_{\beta_0^{(r)}})^{-}n^{-1/2}U(\beta_0^{(r)})+o_P(1). \tag{3.10.40}$$

现在考虑估计量 $\hat{\beta}$. 简单计算可得

$$\frac{2\sqrt{1-||\beta_0^{(r)}||^2}}{\sqrt{1-||\hat{\beta}^{(r)}||^2}+\sqrt{1-||\beta_0^{(r)}||^2}}-1=\frac{\sqrt{1-||\beta_0^{(r)}||^2}-\sqrt{1-||\hat{\beta}^{(r)}||^2}}{\sqrt{1-||\hat{\beta}^{(r)}||^2}+\sqrt{1-||\beta_0^{(r)}||^2}}=O_P(n^{-1/2}).$$

因此

$$\begin{aligned}\sqrt{1-||\hat{\beta}^{(r)}||^2}-\sqrt{1-||\beta_0^{(r)}||^2}&=-\frac{(\hat{\beta}^{(r)}+\beta_0^{(r)})^{\mathrm{T}}(\hat{\beta}^{(r)}-\beta_0^{(r)})}{\sqrt{1-||\hat{\beta}^{(r)}||^2}+\sqrt{1-||\beta_0^{(r)}||^2}}\\&=-\frac{2\beta_0^{(r)\mathrm{T}}(\hat{\beta}^{(r)}-\beta_0^{(r)})+||\hat{\beta}^{(r)}-\beta_0^{(r)}||^2}{\sqrt{1-||\hat{\beta}^{(r)}||^2}+\sqrt{1-||\beta_0^{(r)}||^2}}\\&=-\frac{\beta_0^{(r)\mathrm{T}}(\hat{\beta}^{(r)}-\beta_0^{(r)})}{\sqrt{1-||\beta_0^{(r)}||^2}}+O_P(n^{-1}).\end{aligned}$$

从上面等式可以推出

$$\hat{\beta}-\beta_0=\begin{pmatrix}\hat{\beta}_1-\beta_{01}\\ \vdots\\ \hat{\beta}_{r-1}-\beta_{0(r-1)}\\ -\dfrac{\beta_0^{(r)\mathrm{T}}(\hat{\beta}^{(r)}-\beta_0^{(r)})}{\sqrt{1-||\beta_0^{(r)}||^2}}\\ \hat{\beta}_{r+1}-\beta_{0(r+1)}\\ \vdots\\ \hat{\beta}_p-\beta_{0p}\end{pmatrix}+O_P(n^{-1}).$$

因此

$$\hat{\beta}-\beta_0=J_{\beta_0^{(r)}}(\hat{\beta}^{(r)}-\beta_0^{(r)})+O_P(n^{-1}).\tag{3.10.41}$$

由 (3.10.40) 和 (3.10.41) 式可得

$$\sqrt{n}(\hat{\beta}-\beta_0)=J_{\beta_0^{(r)}}\left(J_{\beta_0^{(r)}}^{\mathrm{T}}DJ_{\beta_0^{(r)}}\right)^{-}n^{-1/2}U(\beta_0^{(r)})+o_P(1).$$

由上式并利用中心极限定理即可证得定理 3.10.2.

推论 3.10.1 的证 由定理 3.10.2 可以推出 $||\hat{\beta}-\beta_0||=O_P(n^{-1/2})$. 因此

$$|(|\hat{\beta}^{\mathrm{T}}\beta_0|-1)|\leqslant|(\hat{\beta}-\beta_0)^{\mathrm{T}}\beta_0|\leqslant||\hat{\beta}-\beta_0||=O_P(n^{-1/2}).$$

这就完成了推论 3.10.1 的证明.

定理 3.10.3 的证　首先证明 (i). 记矩阵 $A = J^{\mathrm{T}}_{\beta_0^{(r)}} D J_{\beta_0^{(r)}}$. 由于 A 是正定矩阵, 因此是可逆的, 记其逆矩阵为 A^{-1}. 定义矩阵

$$\Pi = J_{\beta_0^{(r)}} A^{-1} J^{\mathrm{T}}_{\beta_0^{(r)}}.$$

现在证明 Π 是 D 的广义逆. 为此, 需要证明 $\Pi D \Pi = \Pi$, 且 $D\Pi D = D$. 可以推出

$$\Pi D \Pi = J_{\beta_0^{(r)}} A^{-1} J^{\mathrm{T}}_{\beta_0^{(r)}} D J_{\beta_0^{(r)}} A^{-1} J^{\mathrm{T}}_{\beta_0^{(r)}} = J_{\beta_0^{(r)}} A^{-1} A A^{-1} J^{\mathrm{T}}_{\beta_0^{(r)}} = \Pi.$$

下面证明 $D\Pi D = D$. 首先, 对 β_0 作 QR 分解 (Gentle, 1998, §3.2.2, P95~97), 可以找到它的正交补使得 $B = (b_1, \beta_0)$ 是一个正交矩阵, 且 $\beta_0 = B\begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}$. 因此, $J_{\beta_0^{(r)}} = BB^{\mathrm{T}} J_{\beta_0^{(r)}} \equiv BR$, 其中 $R = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}$, R_1 为 $(p-1)\times(p-1)$ 正定矩阵. 注意到

$$\begin{aligned} A &= J^{\mathrm{T}}_{\beta_0^{(r)}} D J_{\beta_0^{(r)}} = R^{\mathrm{T}} B^{\mathrm{T}} D B R \\ &= R^{\mathrm{T}} \begin{pmatrix} b_1^{\mathrm{T}} D b_1 & 0 \\ 0 & 0 \end{pmatrix} R \\ &= R_1^{\mathrm{T}} b_1^{\mathrm{T}} D b_1 R_1. \end{aligned}$$

因此, 为证 $D\Pi D = D$, 我们将 D 重写为另一种形式. 定义

$$S = B \begin{pmatrix} R_1 & 0 \\ 0 & 1 \end{pmatrix},$$

这里 S 是一个正定矩阵. 那么

$$\begin{aligned} D &= (S^{\mathrm{T}})^{-1} S^{\mathrm{T}} D S S^{-1} \\ &= (S^{\mathrm{T}})^{-1} \begin{pmatrix} R_1^{\mathrm{T}} & 0 \\ 0 & 1 \end{pmatrix} B^{\mathrm{T}} D B \begin{pmatrix} R_1 & 0 \\ 0 & 1 \end{pmatrix} S^{-1} \\ &= (S^{\mathrm{T}})^{-1} \begin{pmatrix} R_1^{\mathrm{T}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b_1^{\mathrm{T}} D b_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} R_1 & 0 \\ 0 & 1 \end{pmatrix} S^{-1} \\ &= (S^{\mathrm{T}})^{-1} \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} S^{-1}. \end{aligned}$$

现在证明 $D\Pi D = D$. 从上面的推导并注意到

$$S^{-1} = \begin{pmatrix} R_1^{-1} & 0 \\ 0 & 1 \end{pmatrix} B^{\mathrm{T}}, \quad (S^{\mathrm{T}})^{-1} = B \begin{pmatrix} (R_1^{\mathrm{T}})^{-1} & 0 \\ 0 & 1 \end{pmatrix},$$

可得

$$
\begin{aligned}
&D\Pi D\\
&=(S^{\mathrm{T}})^{-1}\begin{pmatrix}A&0\\0&0\end{pmatrix}S^{-1}BRQ^{-1}R^{\mathrm{T}}B^{\mathrm{T}}(S^{\mathrm{T}})^{-1}\begin{pmatrix}A&0\\0&0\end{pmatrix}S^{-1}\\
&=(S^{\mathrm{T}})^{-1}\begin{pmatrix}A&0\\0&0\end{pmatrix}\begin{pmatrix}R_1^{-1}&0\\0&1\end{pmatrix}B^{\mathrm{T}}BRA^{-1}R^{\mathrm{T}}B^{\mathrm{T}}B\\
&\quad\times\begin{pmatrix}(R_1^{\mathrm{T}})^{-1}&0\\0&1\end{pmatrix}\begin{pmatrix}A&0\\0&0\end{pmatrix}S^{-1}\\
&=(S^{\mathrm{T}})^{-1}\begin{pmatrix}A&0\\0&0\end{pmatrix}\begin{pmatrix}1\\R_2\end{pmatrix}A^{-1}\begin{pmatrix}1&R_2\end{pmatrix}\begin{pmatrix}A&0\\0&0\end{pmatrix}S^{-1}\\
&=(S^{\mathrm{T}})^{-1}\begin{pmatrix}A\\0\end{pmatrix}A^{-1}\begin{pmatrix}A&0\end{pmatrix}S^{-1}=(S^{\mathrm{T}})^{-1}\begin{pmatrix}A&0\\0&0\end{pmatrix}S^{-1}=D.
\end{aligned}
$$

因此, Π 是 D 的一个广义逆.

下面证明 (ii). 分 $\det(D_{22})\neq 0$ 和 $\det(D_{22})=0$ 两种情况研究. 首先考虑 $\det(D_{22})\neq 0$ 的情况. 记 $J=J_{\beta_0^{(1)}}$, $\alpha=-\beta^{(1)}/\beta_1$. 因为 $\mathrm{rank}(D)=p-1$, $\det(D_{11}D_{22}-D_{12}D_{21})=0$. 注意到 D_{22} 是正定的, 容易证明 $D_{11}=D_{12}D_{22}^{-1}D_{21}$. 另一方面, 可以得到

$$
\begin{aligned}
J^{\mathrm{T}}DJ&=(\alpha\quad I_{p-1})\begin{pmatrix}D_{11}&D_{12}\\D_{21}&D_{22}\end{pmatrix}\begin{pmatrix}\alpha^{\mathrm{T}}\\I_{p-1}\end{pmatrix}\\
&=D_{22.1}+(\sqrt{D_{11}}\alpha+D_{21}/\sqrt{D_{11}})(\sqrt{D_{11}}\alpha^{\mathrm{T}}+D_{12}/\sqrt{D_{11}}),
\end{aligned}\tag{3.10.42}
$$

其中 $D_{22.1}=D_{22}-D_{21}D_{12}/D_{11}$. 结合上式可得 $J^{\mathrm{T}}DJ$ 可逆. 在这种情况下, $(J^{\mathrm{T}}DJ)^-$ 恰好是通常的逆 $(J^{\mathrm{T}}DJ)^{-1}$, 故由 (i) 可得 $J(J^{\mathrm{T}}DJ)^{-1}J^{\mathrm{T}}=D^-$.

当 $\det(D_{22})=0$ 时, 可以得到

$$
D^-=\begin{pmatrix}1/D_{11}+D_{12}D_{22.1}^-D_{21}/D_{11}^2&-D_{12}D_{22.1}^-/D_{11}\\-D_{22.1}^-D_{21}/D_{11}&D_{22.1}^-\end{pmatrix}.
$$

另外, 容易得到

$$
J(J^{\mathrm{T}}DJ)^-J^{\mathrm{T}}=\begin{pmatrix}\alpha^{\mathrm{T}}(J^{\mathrm{T}}DJ)^-\alpha&\alpha^{\mathrm{T}}(J^{\mathrm{T}}DJ)^-\\(J^{\mathrm{T}}DJ)^-\alpha&(J^{\mathrm{T}}DJ)^-\end{pmatrix}.
$$

由 (3.10.42) 式可知 $J^{\mathrm{T}}DJ\geqslant D_{22.1}$. 该不等式与 $\mathrm{rank}(D_{22})=p-2$ 结合可得 $(J^{\mathrm{T}}DJ)^-\leqslant D_{22.1}^-$. 容易验证 $\alpha^{\mathrm{T}}D_{22.1}=0$, 因此 $\alpha\perp\mathrm{span}(D_{22.1})$, $\alpha^{\mathrm{T}}D_{22.1}^-\alpha=0$, 由

此得到 $\alpha^{\mathrm{T}}(J^{\mathrm{T}}DJ)^{-}=0$. 在这种情况下, $J(J^{\mathrm{T}}DJ)^{-}J^{\mathrm{T}}\leqslant D^{-}$, 且有严格的不等号成立, 这是因为 $J^{\mathrm{T}}DJ\neq D_{22.1}$ 且 $1/D_{11}>0$. 这就完成了定理 3.10.3 的证明.

定理 3.10.4 的证 利用引理 3.10.1～ 引理 3.10.3, 可以证得

$$\hat{g}(u;\hat{\beta})-g(u)=\frac{1}{nf(u)}\sum_{i=1}^{n}\varepsilon_iK_h(\beta_0^{\mathrm{T}}X_i-u)+\frac{1}{2}h^2\kappa_2g''(u)+o_P(n^{-1/2}),$$

因此, 由中心极限定理可得即可证得定理 3.10.4.

定理 3.10.5 的证 将 $\hat{\sigma}^2$ 分解可得

$$\begin{aligned}\hat{\sigma}^2&=\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2+\frac{1}{n}\sum_{i=1}^{n}[g(\beta_0^{\mathrm{T}}X_i)-\hat{g}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})]^2\\&\quad+\frac{2}{n}\sum_{i=1}^{n}\varepsilon_i[g(\beta_0^{\mathrm{T}}X_i)-\hat{g}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})]\\&\equiv I_1+I_2+I_3.\end{aligned}$$

利用定理 3.10.1 可得

$$\begin{aligned}\sqrt{n}|I_2|&\leqslant\sqrt{n}\sup_{u\in\mathcal{U},\beta\in\mathcal{B}_n}|\hat{g}(u;\beta)-g(u)|^2\\&=O_P((nh^2/\log^2 n)^{-1/2})\xrightarrow{P}0.\end{aligned}$$

注意到 $E(\varepsilon_i)=0$, 类似于 (3.10.35) 式的证明, 可以证明 $\sqrt{n}I_3\xrightarrow{P}0$. 这就证明了 $\hat{\sigma}^2=n^{-1}\sum_{i=1}^{n}\varepsilon_i^2+o_P(n^{-1/2})$. 因此

$$\sqrt{n}(\hat{\sigma}^2-\sigma^2)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\varepsilon_i^2-\sigma^2)+o_P(1).$$

对上式利用中心极限定理即可证得定理 3.10.5.

3.11 拟似然估计和估计函数方法估计

本节考虑异方差结构的单指标模型

$$E(Y_i|X_i)=g(\beta_0^{\mathrm{T}}X_i),\quad \mathrm{var}(Y_i|X_i)=\sigma^2V(g(\beta_0^{\mathrm{T}}X_i)),\tag{3.11.1}$$

其中 $V(\cdot)$ 是已知的方差函数, $g(\cdot)$ 是未知的单变量联系函数, σ^2 是未知参数, β_0 是未知的指标系数. 通常假定 $||\beta_0||=1$, 并且 β_0 的第一个非零元素大于 0; 这个假定是为了模型的可识别性.

如果 $g(\cdot)$ 是已知的, 那么对数拟似然比可定义为

$$Q(\beta)=\sum_{i=1}^{n}\int_{Y_i}^{g(\beta^{\mathrm{T}}X_i)}\frac{Y_i-u}{V(u)}\mathrm{d}u.$$

极大化 $Q(\beta)$ 可以得到 β_0 的极大拟似然估计. 这个估计量等价于方程组

$$\begin{cases}\displaystyle\sum_{i=1}^{n}\frac{g'(\beta^{\mathrm{T}}X_i)}{V(g(\beta^{\mathrm{T}}X_i))}J_{\beta^{(r)}}^{\mathrm{T}}X_i[Y_i-g(\beta^{\mathrm{T}}X_i)]=0,\\ ||\beta||-1=0\end{cases}$$

的解, 其中 $J_{\beta^{(r)}}$ 由 (3.10.2) 式定义. 然而, $g(\cdot)$ 和 $g'(\cdot)$ 是未知的, 它们可以分别用其局部线性估计 $\hat{g}(\cdot)$ 和 $\hat{g}'(\cdot)$ 代替, 见 (3.10.4) 和 (3.10.5) 式. 因而可得方程组

$$\begin{cases}\displaystyle\sum_{i=1}^{n}\frac{\hat{g}'(\beta^{\mathrm{T}}X_i;\beta)}{V(\hat{g}(\beta^{\mathrm{T}}X_i;\beta))}J_{\beta^{(r)}}^{\mathrm{T}}X_i[Y_i-\hat{g}(\beta^{\mathrm{T}}X_i;\beta)]=0,\\ ||\beta||-1=0.\end{cases}\tag{3.11.2}$$

解这个方程组即可得到 β_0 的极大拟似然估计, 并记为 $\hat{\beta}_{\mathrm{QL}}$.

Chiou 和 Müller(1998) 在固定设计点列下证明了极大拟似然估计的渐近正态性, 并证明了它是渐近有效的. 按照定理 3.10.2 的证明方法可以证明估计量 $\hat{\beta}_{\mathrm{QE}}$ 的渐近正态, 即在适当的正则条件下, 有

$$\sqrt{n}(\hat{\beta}_{\mathrm{QL}}-\beta_0)\xrightarrow{D}N\left(0,\ \sigma^2 J_{\beta_0^{(r)}}(J_{\beta_0^{(r)}}^{\mathrm{T}}\widetilde{D}J_{\beta_0^{(r)}})^{-}J_{\beta_0^{(r)}}^{\mathrm{T}}\right).\tag{3.11.3}$$

其中

$$\widetilde{D}=E\left\{\frac{g'(\beta_0^{\mathrm{T}}X)^2}{V(g(\beta_0^{\mathrm{T}}X))}[X-E(X|\beta_0^{\mathrm{T}}X)][X-E(X|\beta_0^{\mathrm{T}}X)]^{\mathrm{T}}\right\}.\tag{3.11.4}$$

在同方差的情况下, 即 $V(\cdot)\equiv 1$, 逆似然估计 $\hat{\beta}_{\mathrm{QL}}$ 即成为 3.10 节的估计方程估计.

需要指出的是: 在研究由方程组 (3.11.2) 得到的估计量的渐近正态性时, 需要对联系函数 $g(\cdot)$ 进行欠光滑, 这就为选择带宽带来了困难. 为此, Cui 等 (2011) 利用纠偏技巧将估计方程组 (3.11.2) 进行了修正, 即考虑了如下方程组:

$$\begin{cases}R_n^*(\beta^{(r)})=0,\\ ||\beta||-1=0,\end{cases}\tag{3.11.5}$$

其中

$$R_n(\beta^{(r)})=\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{g}'(\beta^{\mathrm{T}}X_i;\beta)}{V(\hat{g}(\beta^{\mathrm{T}}X_i;\beta))}J_{\beta^{(r)}}^{\mathrm{T}}[X_i-\hat{\mu}(\beta^{\mathrm{T}}X_i;\beta)][Y_i-\hat{g}(\beta^{\mathrm{T}}X_i;\beta)],$$

$\hat{\mu}(\cdot;\beta)$ 在 (3.10.10) 式中定义. 解方程组 (3.11.5) 即可得到 β_0 的估计, 记为 $\hat{\beta}_{\rm EFM}$, 称之为估计函数方法 (EFM) 估计.

可以用迭代公式

$$\hat{\beta} = \tilde{\beta} + J_{\tilde{\beta}^{(r)}} B_n^{-}(\tilde{\beta}^{(r)}) R_n(\tilde{\beta}^{(r)}). \tag{3.11.6}$$

得到方程组 (3.11.5) 的近似解, 其中

$$B_n(\tilde{\beta}^{(r)}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{g}'(\tilde{\beta}^{\rm T}X_i;\beta)^2}{V(\hat{g}(\tilde{\beta}^{\rm T}X_i;\beta))} J_{\tilde{\beta}^{(r)}}^{\rm T} [X_i - \hat{\mu}(\tilde{\beta}^{\rm T}X_i;\tilde{\beta})]^{\otimes 2} J_{\tilde{\beta}^{(r)}},$$

$A^{\otimes 2} = AA^{\rm T}$, A 为任意矩阵.

Cui 等 (2011) 给出了另一个迭代公式. 下面我们叙述这个迭代公式的推导过程. 对任何参数向量 β, 记 $\hat{F}(\beta) = (\hat{F}_1(\beta), \cdots, \hat{F}_p(\beta))^{\rm T}$, 其中

$$\hat{F}_s(\beta) = \sum_{i=1}^{n} \frac{\hat{g}'(\beta^{\rm T}X_i;\beta)}{V(\hat{g}(\beta^{\rm T}X_i;\beta))} [X_{si} - \hat{\mu}_s(\beta^{\rm T}X_i;\beta)][Y_i - \hat{g}(\beta^{\rm T}X_i;\beta)].$$

这里 X_{si} 和 $\hat{\mu}_s$ 分别为 X_i 和 $\hat{\mu}$ 的第 s 个分量, $s = 1, \cdots, p$. 为表达方便起见, 不妨设 β 的第一个分量大于 0. 因此, 方程组 (3.11.5) 可写为

$$\begin{cases} -\beta_2 \hat{F}_1(\beta)/\sqrt{1-||\beta^{(1)}||^2} + \hat{F}_2(\beta) = 0, \\ \qquad\cdots\cdots \\ -\beta_p \hat{F}_1(\beta)/\sqrt{1-||\beta^{(1)}||^2} + \hat{F}_p(\beta) = 0, \\ ||\beta|| - 1 = 0. \end{cases}$$

注意到 $\beta_1 = \sqrt{1-||\beta^{(1)}||^2}$, 简单计算可得

$$\begin{cases} \beta_1 = |\hat{F}_1(\beta)|/||\hat{F}(\beta)||, \\ \beta_s^2 = \hat{F}_s^2(\beta)/||\hat{F}(\beta)||^2, \quad s \geqslant 2, \end{cases}$$

且 $\operatorname{sgn}\{\beta_s \hat{F}_1(\beta)\} = \operatorname{sgn}\{\hat{F}_s(\beta)\}$, $s \geqslant 2$. 上面的方程组也可以写为

$$\frac{\hat{F}_1(\beta)}{||\hat{F}(\beta)||}\beta = \frac{|\hat{F}_1(\beta)|}{||\hat{F}(\beta)||} \cdot \frac{\hat{F}(\beta)}{||\hat{F}(\beta)||}. \tag{3.11.7}$$

因此, 解方程组 (3.11.5) 等价于寻找 (3.11.7) 式的不动点. 虽然在 (3.11.7) 式中 $||\beta^{(1)}|| < 1$ 几乎处处成立, 且总有 $||\beta|| = 1$, 但 (3.11.7) 式直接用来作为迭代方程将会引起一些麻烦. 注意到 $||\hat{F}(\beta)||$ 的值用作分母有时可能很小, 会潜在地出现算法不稳定的情况. 另一方面, 由 (3.11.7) 式得到的不动点迭代算法的收敛速度依赖于

L, 其中 $\left\|\dfrac{\partial\{\hat{F}(\beta)/\|\hat{F}(\beta)\|\}}{\partial\beta}\right\| \leqslant L$. 为得到快的收敛速度, 技术上需要给一个门限值 L. 引入一个专门固定的常数 M, 将 (3.11.7) 式两边加 $M\beta$, 并用 $\hat{F}_1(\beta)/\|\hat{F}(\beta)\|+M$ 除可得

$$\beta = \frac{M}{\hat{F}_1(\beta)/\|\hat{F}(\beta)\|+M}\beta + \frac{|\hat{F}_1(\beta)|/\|\hat{F}(\beta)\|^2}{\hat{F}_1(\beta)/\|\hat{F}(\beta)\|+M}\hat{F}(\beta), \tag{3.11.8}$$

其中选择 M 使得 $\hat{F}_1(\beta)/\|\hat{F}(\beta)\|+M \neq 0$. 另外, 为加快收敛速度, 我们通过选择某个适当的常数 M 来减少上式右边的衍生项. 这就是下述迭代程序的思想.

基于上述讨论, 不动点迭代算法概述如下.

步骤 0. 选择 β_0 的一个模为 1 的初始值, 并记为 β_{old}(初始化步).

步骤 1. 利用公式 (3.10.4) 和 (3.10.5) 计算局部线性估计 $\hat{g}(\beta_{\text{old}}^{\mathrm{T}}X_i;\beta_{\text{old}})$ 和 $\hat{g}'(\beta_{\text{old}}^{\mathrm{T}}X_i;\beta_{\text{old}})$, $i=1,\cdots,n$. 这里估计量 $\hat{g}$ 和 $\hat{g}'$ 使用同一个带宽.

步骤 2. 使用公式 (3.11.8) 计算 β 的一个新的值, 并记为 β_{new}, 即

$$\beta_{\text{new}} = \frac{M}{\hat{F}_1(\beta_{\text{old}})/\|\hat{F}(\beta_{\text{old}})\|+M}\beta_{\text{old}} + \frac{|\hat{F}_1(\beta_{\text{old}})|/\|\hat{F}(\beta_{\text{old}})\|^2}{\hat{F}_1(\beta_{\text{old}})/\|\hat{F}(\beta_{\text{old}})\|+M}\hat{F}(\beta_{\text{old}}),$$

其中常数 M 满足: 对任何 β, $\hat{F}_1(\beta)/\|\hat{F}(\beta)\|+M \neq 0$; 然后更新 β_{old}, 即设置 $\beta_{\text{old}} = \beta_{\text{new}}/\|\beta_{\text{new}}\|$.

步骤 3. 重复步骤 1 和 2 直到 $\|\beta_{\text{new}}-\beta_{\text{old}}\| \leqslant \text{tol}$, 这里 tol 是一个规定的偏差. 最后迭代得到的向量 $\beta_{\text{new}}/\|\beta_{\text{new}}\|$ 即为 β_0 的估计值.

在步骤 0 中, 初始值以经验可取为 $(1,\cdots,1)^{\mathrm{T}}/\sqrt{p}$. 在步骤 1 中使用的带宽可选为基于观察 $\{\beta_{\text{old}}^{\mathrm{T}}X_i, Y_i\}$ 得到的估计 $\hat{g}$ 的最优带宽. 因此可以采用诸如 K- 折叠交叉验证和广义交叉验证来确定最优带宽. 在步骤 2 中, β_{new} 的模不等于 1, 我们必须再对它作标准化处理. 因为步骤 2 中的迭代导致 β_{new} 违反了模 1 的可识别性限制, 所以我们设计 (3.11.7) 式包含全部的 β 向量. 在步骤 2 的每一步迭代中对 β_{new} 再标准化而避开了控制 $\|\beta_{\text{new}}^{(1)}\| < 1$ 的困难. 最后指出, 适当的选取调整因子 M 将提高迭代的稳定性. 在每一步迭代中选择 M 的最优值以保证步骤 2 的迭代公式的右边的衍生项接近于 0. 按照这个思想, 如果 M 的值随着 β 和 $\hat{F}(\beta)/\|\hat{F}(\beta)\|$ 的变化而改变. 那么计算量将会很大. 因此, 我们在每一步迭代中希望 M 是一个不变的常数, 并按照极小化模型预测误差并用 K- 折叠交叉验证法选择 M; 当维数 p 变的更大时, M 将变的更小. 通常作模拟计算时在区间 $[2/\sqrt{p}, p/2]$ 内以经验确定 M 的值.

Cui 等 (2011) 证明了 EFM 估计的渐近正态性. 即在适当的正则条件下, 估计量 $\hat{\beta}_{\text{EFM}}$ 具有 (3.11.3) 式给出的渐近正态性. 正如定理 3.10.3 给出的结果, 这个结果也说明 EFM 估计比 Härdle 等 (1993) 给出的估计有效.

需要指出的是, 由拟似然法和 EFM 得到的估计量有更小的渐近方差得益于再参数化方法. 对照大多数对 β_0 现有的估计方法, 似乎都是在线性降维框架下来做的. 线性降维的目的是寻找能将原向量变换成少一维向量的方向, 如平均导数估计和 SIR 是两个与之相关的方法.

3.12 经验似然推断

本节主要考虑模型 (3.1.1) 的经验似然推断问题, 提出指标系数的几种经验似然比函数的构造方法, 解决指标系数的置信域的构造问题, 同时也作模拟研究. 本节的主要内容取材于 (Xue and Zhu, 2006; Zhu and Xue, 2006).

3.12.1 估计的经验似然

为利用 $\|\beta_0\|=1$ 提供的信息, 我们仍采用 3.10.1 小节中 "去一分量" 的方法对 β 再参数化, 可知 $\beta=\beta(\beta^{(r)})$ 为 $\beta^{(r)}$ 的函数, 其中 $\beta^{(r)}=(\beta_1,\cdots,\beta_{r-1},\beta_{r+1},\cdots,\beta_p)^{\mathrm{T}}$, 并得到 β 关于 $\beta^{(r)}$ 的 Jacobi 矩阵 (3.10.3). 因此, 当 $g(\cdot)$ 已知时, 对任何单位向量 β, 引入一个辅助随机变量

$$Z_i(\beta^{(r)})=[Y_i-g(\beta^{\mathrm{T}}X_i)]g'(\beta^{\mathrm{T}}X_i)J_{\beta^{(r)}}^{\mathrm{T}}X_i.$$

注意到当 $\beta=\beta_0$ 时, $E[Z_i(\beta_0^{(r)})]=0$. 反之, 当 $E[Z_i(\beta^{(r)})]=0$ 时, 我们可以构造一个估计方程 $\sum\limits_{i=1}^{n}Z_i(\beta^{(r)})=0$, 它的解恰好是 $\beta^{(r)}$ 的最小二乘估计. 因此, 按照 Owen(1988, 1990) 的思想, 可以定义 profile 经验对数似然比函数

$$l_n(\beta^{(r)})=-2\max\left\{\sum_{i=1}^{n}\log(np_i)\bigg|\,p_i\geqslant 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_iZ_i(\beta^{(r)})=0\right\}.$$

可以证明: $l_n(\beta_0^{(r)})$ 以分布收敛于自由度为 $p-1$ 的 χ^2 变量. 然而, $l_n(\beta^{(r)})$ 不能被直接用来对 $\beta_0^{(r)}$ 作统计推断, 因为 $l_n(\beta^{(r)})$ 包含两个未知量 $g(\cdot)$ 和 $g'(\cdot)$. 解决这个问题的一个自然的方法是用 $g(\cdot)$ 和 $g'(\cdot)$ 的估计量来代替. 这里使用局部线性估计. 他们已在 (3.10.4) 和 (3.10.5) 中定义.

用 $\hat{g}(\beta^{\mathrm{T}}X_i;\beta)$ 和 $\hat{g}'(\beta^{\mathrm{T}}X_i;\beta)$ 分别代替 $g(\beta^{\mathrm{T}}X_i)$ 和 $g'(\beta^{\mathrm{T}}X_i)$ 可得到 $Z_i(\beta^{(r)})$ 的估计, 并记为设 $\hat{Z}_i(\beta^{(r)})$, $i=1,\cdots,n$. 那么, 一个估计的经验对数似然比函数定义为

$$\hat{l}_{\mathrm{EEL}}(\beta^{(r)})=-2\max\left\{\sum_{i=1}^{n}\log(np_i)\bigg|\,p_i\geqslant 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\hat{Z}_i(\beta^{(r)})=0\right\}.$$

利用 Lagrange 乘子法, $\hat{l}_{\text{EEL}}(\beta^{(r)})$ 可表示为

$$\hat{l}_{\text{EEL}}(\beta^{(r)}) = 2\sum_{i=1}^{n}\log\big(1+\lambda^{\text{T}}\hat{Z}_i(\beta^{(r)})\big), \tag{3.12.1}$$

其中 λ 由下式确定:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Z}_i(\beta^{(r)})}{1+\lambda^{\text{T}}\hat{Z}_i(\beta^{(r)})} = 0. \tag{3.12.2}$$

为给出本节的主要结果, 除条件 (C3.10.1)~(C3.10.4) 外, 还需要下列条件.

(C3.12.1) $nh^2 \to \infty$, $nh^4 \to 0$, $nhh_1^3 \to \infty$, $nh_1^5 = O(1)$.

(C3.12.2) $V_0(\beta_0^{(r)})$ 和 $V(\beta_0^{(r)})$ 为正定矩阵, 其中

$$V_0(\beta_0^{(r)}) = E[\varepsilon^2 g'(\beta_0^{\text{T}}X)^2 J_{\beta_0^{(r)}}^{\text{T}} XX^{\text{T}} J_{\beta_0^{(r)}}],$$

$$V(\beta_0^{(r)}) = E\{\varepsilon^2 g'(\beta_0^{\text{T}}X)^2 J_{\beta_0^{(r)}}^{\text{T}}[X-E(X|\beta_0^{\text{T}}X)][X-E(X|\beta_0^{\text{T}}X)]^{\text{T}} J_{\beta_0^{(r)}}\}.$$

Xue 和 Zhu(2006) 证明了估计的经验似然比的渐近性质, 即下面的定理.

定理 3.12.1 设条件 (C3.12.1)、(C3.12.2) 和 (C3.10.1)~(C3.10.4)(i) 成立, 则

$$\hat{l}_{\text{EEL}}(\beta_0^{(r)}) \xrightarrow{D} w_1\chi_{1,1}^2 + \cdots + w_{p-1}\chi_{1,p-1}^2,$$

其中 $w_i(1 \leqslant i \leqslant p-1)$ 是矩阵 $D(\beta_0^{(r)}) = V_0^{-1}(\beta_0^{(r)})V(\beta_0^{(r)})$ 的特征根, $\chi_{1,i}^2(1 \leqslant i \leqslant p-1)$ 是自由度为 1 的独立 χ^2 变量.

为应用定理 3.12.1 构造 $\beta_0^{(r)}$ 的置信域, 我们需要给出未知权 w_i 的相合估计. 用 "嵌入" 法可得 $V_0(\beta_0^{(r)})$ 和 $V(\beta_0^{(r)})$ 的估计, 即

$$\hat{V}_0(\hat{\beta}^{(r)}) = \frac{1}{n}\sum_{i=1}^{n}\{[Y_i-\hat{g}(\hat{\beta}^{\text{T}}X_i;\hat{\beta})]^2\hat{g}'(\hat{\beta}^{\text{T}}X_i;\hat{\beta})^2 J_{\hat{\beta}^{(r)}}^{\text{T}} X_iX_i^{\text{T}} J_{\hat{\beta}^{(r)}}\},$$

$$\begin{aligned}\hat{V}(\hat{\beta}^{(r)}) = {} & \frac{1}{n}\sum_{i=1}^{n}\{[Y_i-\hat{g}(\hat{\beta}^{\text{T}}X_i;\hat{\beta})]^2\hat{g}'(\hat{\beta}^{\text{T}}X_i;\hat{\beta})^2 \\ & \times J_{\hat{\beta}^{(r)}}^{\text{T}}[X_i-\hat{\mu}(\hat{\beta}^{\text{T}}X_i;\hat{\beta})][X_i-\hat{\mu}(\hat{\beta}^{\text{T}}X_i;\hat{\beta})]^{\text{T}} J_{\hat{\beta}^{(r)}}\},\end{aligned} \tag{3.12.3}$$

其中 $\hat{\beta} = (\hat{\beta}_1, \cdots, \hat{\beta}_p)^{\text{T}}$ 是 β_0 的估计方程估计 (见 3.10.1 小节), $\hat{\beta}^{(r)} = (\hat{\beta}_1, \cdots, \hat{\beta}_{r-1}, \hat{\beta}_{r+1}, \cdots, \hat{\beta}_p)^{\text{T}}$, $\hat{\mu}(\cdot;\hat{\beta})$ 是由 (3.10.10) 定义的 $\mu(\cdot) = E(X|\hat{\beta}^{\text{T}}X = \cdot)$ 的估计. 这就推出 $\hat{D}(\hat{\beta}^{(r)}) = \hat{V}_0^{-1}(\hat{\beta}^{(r)})\hat{V}(\hat{\beta}^{(r)})$ 的特征根 $\hat{w}_i$ 是 w_i 的相合估计, $1 \leqslant i \leqslant p-1$. 设

$H(\cdot)$ 为给定样本 $\{(X_i, Y_i), 1 \leqslant i \leqslant n\}$ 下加权和 $\hat{s} = \hat{w}_1\chi^2_{1,1} + \cdots + \hat{w}_{p-1}\chi^2_{1,p-1}$ 的条件分布, 并设 $\hat{c}_\alpha$ 是 $H(\cdot)$ 的 $1-\alpha$ 分位数. 那么 $\beta_0^{(r)}$ 的近似 $1-\alpha$ 置信域为

$$\hat{I}_\alpha(\beta^{(r)}) = \{\beta^{(r)} | \hat{l}_{\mathrm{EEL}}(\beta^{(r)}) \leqslant \hat{c}_\alpha, ||\beta^{(r)}|| < 1\}.$$

在实际中, 通过从 χ^2_1 分布中重复产生独立样本 $\chi^2_{1,1}, \cdots, \chi^2_{1,p-1}$, 并用 Monte-Carlo 模拟可以得到 $H(\cdot)$.

3.12.2 调整的经验似然

注意到为得到估计的经验对数似然比, 需要 Monte-Carlo 模拟, 因此计算量很大. 在本节中, 我们介绍两个调整方法以避免极限分布的 Monte-Carlo 模拟计算.

设 $\rho(\beta^{(r)}) = (p-1)/\mathrm{tr}\{D(\beta^{(r)})\}$, 其中 $\mathrm{tr}(\cdot)$ 是迹运算. 那么遵照 Rao 和 Scott (1981), 用自由度为 $p-1$ 的 χ^2 分布 χ^2_{p-1} 可以逼近 $\rho(\beta^{(r)})\sum_{i=1}^{p-1} w_i\chi^2_{1,i}$ 的分布. 这就推出用 χ^2_{p-1} 可以逼近 Rao-Scott 调整的经验对数似然比 $\hat{\rho}(\hat{\beta}^{(r)})\hat{l}_{\mathrm{EEL}}(\beta^{(r)})$. 使用定理 3.12.1 及 $\hat{V}(\hat{\beta}^{(r)})$ 和 $\hat{V}_0(\beta^{(r)})$ 的相合性可以达到这个目的, 这里 $\hat{\rho}(\hat{\beta}^{(r)}) = (p-1)/\mathrm{tr}\{\hat{D}(\hat{\beta}^{(r)})\}$. 通过用 $\beta^{(r)}$ 代替 $\hat{\rho}(\hat{\beta}^{(r)})$ 中的 $\hat{\beta}^{(r)}$ 可以改进 Rao-Scott 调整的经验对数似然比. 因此, 一个改进的 Rao-Scott 调整的经验对数似然比函数定义为

$$\tilde{l}(\beta^{(r)}) = \hat{\rho}(\beta^{(r)})\hat{l}_{\mathrm{EEL}}(\beta^{(r)}).$$

然而, 这个逼近的精度依赖于诸 w_i 的值. 下面给出一个调整的经验对数似然比, 其渐近分布是自由度为 $p-1$ 的标准 χ^2 分布. 注意到

$$\hat{\rho}(\hat{\beta}^{(r)}) = \frac{\mathrm{tr}\{\hat{V}^{-1}(\hat{\beta}^{(r)})\hat{V}(\hat{\beta}^{(r)})\}}{\mathrm{tr}\{\hat{V}_0^{-1}(\hat{\beta}^{(r)})\hat{V}(\hat{\beta}^{(r)})\}}.$$

通过仔细检查 $\hat{l}_{\mathrm{EEL}}(\beta^{(r)})$ 的渐近表达, 我们用

$$\hat{B}(\hat{\beta}^{(r)}) = \left[\sum_{i=1}^{n} \hat{Z}_i(\hat{\beta}^{(r)})\right]\left[\sum_{i=1}^{n} \hat{Z}_i(\hat{\beta}^{(r)})\right]^{\mathrm{T}}$$

代替 $\hat{\rho}(\hat{\beta}^{(r)})$ 中的 $\hat{V}(\hat{\beta}^{(r)})$, 并得到不同的调整因子

$$\hat{r}(\hat{\beta}^{(r)}) = \frac{\mathrm{tr}\{\hat{V}^{-1}(\hat{\beta}^{(r)})\hat{B}(\hat{\beta}^{(r)})\}}{\mathrm{tr}\{\hat{V}_0^{-1}(\hat{\beta}^{(r)})\hat{B}(\hat{\beta}^{(r)})\}}.$$

可以证明 $\hat{r}(\hat{\beta}^{(r)})\hat{l}_{\mathrm{EEL}}(\beta^{(r)})$ 渐近于 χ^2_{p-1} 变量. 为提高逼近的精度, 用 $\beta^{(r)}$ 代替 $\hat{r}(\hat{\beta}^{(r)})$ 中的 $\hat{\beta}^{(r)}$, 并定义一个调整的经验对数似然比, 即

$$\hat{l}_{\mathrm{AEL}}(\beta^{(r)}) = \hat{r}(\beta^{(r)})\hat{l}_{\mathrm{EEL}}(\beta^{(r)}). \tag{3.12.4}$$

Xue 和 Zhu(2006) 证明下述定理.

定理 3.12.2 设条件 (C3.12.1)、(C3.12.2) 和 (C3.10.1)~(C3.10.4)(i) 成立, 则

$$\hat{l}_{\text{AEL}}(\beta_0^{(r)}) \xrightarrow{D} \chi^2_{p-1}.$$

基于定理 3.12.2, 可以使用 $\hat{l}_{\text{AEL}}(\beta^{(r)})$ 构造 $\beta_0^{(r)}$ 的置信域. 设

$$\hat{I}_{\text{AEL},\alpha}(\beta^{(r)}) = \left\{\beta^{(r)} \,|\, \hat{l}_{\text{AEL}}(\beta^{(r)}) \leqslant \chi^2_{p-1}(1-\alpha), ||\beta^{(r)}|| < 1\right\}.$$

那么 $\hat{I}_{\text{AEL},\alpha}(\beta^{(r)})$ 给出了 $\beta_0^{(r)}$ 的近似 $1-\alpha$ 置信域.

3.12.3 纠偏的经验似然

在前面两个方法中, 为了使提出的经验对数似然比渐近于标准 χ^2 变量, 需要对联系函数 $g(\cdot)$ 的估计进行欠光滑, 这就带来了选择带宽的困难. 为解决这一问题, Zhu 和 Xue(2006) 提出了一个 "纠偏" 方法, 使得不需要做欠光滑就可以使经验对数似然比渐近于标准 χ^2 变量.

注意到 (3.12.1) 式中的 $\hat{Z}(\beta^{(r)})$ 是加权残差. 当使用最优带宽时为消除比 $\sqrt{n}$ 更低的收敛速度的余项, 我们考虑将 X 条件中心化来纠正偏差. 具体理由将在注 3.12.1 中给予详细阐述. 为此, 构造辅助随机变量

$$\hat{\eta}_i(\beta^{(r)}) = [Y_i - \hat{g}(\beta^{\text{T}}X_i;\beta)]\hat{g}'(\beta^{\text{T}}X_i;\beta)J^{\text{T}}_{\beta^{(r)}}[X_i - \hat{\mu}(\beta^{\text{T}}X_i;\beta)],$$

其中 $\hat{\mu}(u;\beta)$ 为由 (3.10.10) 式定义的 $\mu(u) = E[X|\beta^{\text{T}}X = u]$ 的估计量. 因此, 一个纠偏的经验对数似然比函数定义为

$$\hat{l}_{\text{BCEL}}(\beta^{(r)}) = -2\max\left\{\sum_{i=1}^n \log(np_i) : p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\hat{\eta}_i(\beta^{(r)}) = 0\right\}.$$

利用 Lagrange 乘子法, 可将 $\hat{l}_{\text{BCEL}}(\beta^{(r)})$ 表示为

$$\hat{l}_{\text{BCEL}}(\beta^{(r)}) = 2\sum_{i=1}^n \log(1 + \lambda^{\text{T}}\hat{\eta}_i(\beta^{(r)})), \tag{3.12.5}$$

其中 λ 由下式确定:

$$\frac{1}{n}\sum_{i=1}^n \frac{\hat{\eta}_i(\beta^{(r)})}{1 + \lambda^{\text{T}}\hat{\eta}_i(\beta^{(r)})} = 0. \tag{3.12.6}$$

注 3.12.1 由 (3.12.5) 和 (3.12.6) 式并利用 Taylor 展式, 可以证明 $\hat{l}_{\text{BCEL}}(\beta_0^{(r)})$ 等价于

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\eta}_i(\beta_0^{(r)})\right)^{\text{T}} \left(\frac{1}{n}\sum_{i=1}^n \hat{\eta}_i(\beta_0^{(r)})\hat{\eta}_i^{\text{T}}(\beta_0^{(r)})\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\eta}_i(\beta_0^{(r)})\right).$$

那么, 为证明这项的渐近 χ^2 性, 需要证明

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\eta}_i(\beta_0^{(r)})-\varepsilon_i g'(\beta_0^{\mathrm{T}}X_i)J_{\beta_0^{(r)}}^{\mathrm{T}}[X_i-E(X_i|\beta_0^{\mathrm{T}}X_i)]\right\}$$

是渐近可忽略的. 当计算这个差的二阶矩时, 因子 $X_i-E(X_i|\beta^{\mathrm{T}}X_i)$ 起着重要作用, 且可以证明这个余项的二阶矩收敛于 0. 此外, 这样的纠偏使带宽的选择范围变为 $(c_1n^{-1/4},c_2n^{-1/7})$, 其中 $c_1>0$ 和 $c_2>0$ 为常数. 这个范围在非参数估计的意义上含盖了最优带宽.

对纠偏的经验似然比, Zhu 和 Xue(2006) 证明了下列结果:

定理 3.12.3　设条件 (C3.12.2) 和 (C3.10.1)~(C3.10.4)(i) 成立. 如果 $nh^3\to\infty$, $nh^8\to 0$, 则

$$\hat{l}_{\mathrm{BCEL}}(\beta_0^{(r)})\xrightarrow{\mathcal{D}}\chi_{p-1}^2.$$

基于定理 3.12.3, 可以构造 $\beta_0^{(r)}$ 的近似 $1-\alpha$ 置信域

$$\hat{I}_{\mathrm{BCEL},\alpha}(\beta^{(r)})=\left\{\beta^{(r)}|\,\hat{l}_{\mathrm{BCEL}}(\beta^{(r)})\leqslant\chi_{p-1}^2(1-\alpha),||\beta^{(r)}||<1\right\}.$$

通过关系式 $\beta_r=(1-||\beta^{(r)}||^2)^{1/2}$, 可由 $\beta_0^{(r)}$ 的置信域得到 β_0 的置信域. 定理 3.12.1~ 定理 3.12.3 也可以用来对 β_0 作大样本假设检验.

3.12.4　模拟研究

本节对 $p=2$, $p=3$ 和 $p=4$ 三种情况进行模拟研究. 为了比较, 使用下列四种方法作模拟运算: 纠偏的经验似然 (BCEL)、调整的经验似然 (AEL)、估计的经验似然 (EEL) 和正态逼近 (NA) 方法. 通过置信区间的长度和覆盖概率的大小进行比较. 正态逼近的置信域是基于定理 3.10.2 给出的 $\hat{\beta}$ 的渐近正态分布, 其估计的渐近方差矩阵为 $\hat{\Sigma}=\hat{\sigma}^2J_{\hat{\beta}^{(r)}}(J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\hat{D}J_{\hat{\beta}^{(r)}})^{-}J_{\hat{\beta}^{(r)}}^{\mathrm{T}}$, 其中 $\hat{\sigma}^2$ 和 $\hat{D}$ 分别由 (3.10.8) 和 (3.10.9) 式定义.

当 $p=2$ 和 $p=3$ 时, 我们可以计算覆盖概率并能画出置信域的图形. 但当 $p=4$ 时, 我们只能报告覆盖概率因为画出图形很困难. 在下列三个模拟例子中, 样本量是 $n=150,200,300$. 核函数取为

$$K(u)=\frac{15}{16}(1-u^2)^2I[|u|\leqslant 1].$$

另一个重要的事情是带宽 h 和 h_1 的选择. 根据条件 (C3.12.1), h 的速度在 $n^{-1/2}$ 和 $n^{-1/4}$ 之间, h_1 的速度范围是 $(nh)^{-1/3}$ 到 $n^{-1/5}$. 注意到由现有方法选择的最优带宽 h_{opt} 具有速度 $n^{-1/5}$, 如交叉验证 (CV) 或广义交叉验证 (GCV). 因此, h 和 h_1 可以分别选择为

$$h=h_{\mathrm{opt}}n^{-2/15}\quad 和\quad h_1=h_{\mathrm{opt}}.$$

由上式可知 $h=O(n^{-1/3})$ 和 $h_1=O(n^{-1/5})$. 二者都满足条件 (C3.12.1). 相关的讨论可参看 (Carroll et al., 1997; Stute and Zhu, 2005; Chiou and Müller, 1998). 下面的模拟使用 CV 方法.

通过 500 次模拟运行计算近似 0.95 置信域和他们的覆盖概率, 对于 EEL, 重抽样的样本量是 500.

例3.12.1 考虑单指标模型

$$Y=15\exp\{-(\beta_1X_1+\beta_2X_2)^2\}+\varepsilon,$$

其中 $(\beta_1,\beta_2)=(0.8,0.6)$, ε 服从标准正态分布, X_1 和 X_2 是相互独立的随机变量, 且具有联合密度

$$f(x_1,x_2)=\frac{1}{2\pi}\exp\left(\frac{x_1^2+x_2^2}{2}\right)I_{[-4,4]}(x_1)I_{[-4,4]}(x_2).$$

模拟结果报告在表 3.12.1 中.

表 3.12.1 当维数 $p=2$ 时, 在不同的样本量 n 下 β_1 和 β_2 的近似 0.95 置信区间的平均长度和经验覆盖概率

β	n	平均长度				覆盖概率			
		BCEL	AEL	EEL	NA	BCEL	AEL	EEL	NA
β_1	150	0.0229	0.0229	0.0226	0.0205	0.920	0.916	0.910	0.890
	200	0.0196	0.0196	0.0195	0.0160	0.928	0.924	0.924	0.898
	300	0.0155	0.0155	0.0154	0.0148	0.944	0.940	0.938	0.914
β_2	150	0.0306	0.0306	0.0302	0.0274	0.920	0.916	0.910	0.890
	200	0.0261	0.0261	0.0260	0.0240	0.926	0.924	0.924	0.898
	300	0.0207	0.0207	0.0206	0.0197	0.944	0.940	0.938	0.918

从表 3.12.1 的结果可以看出 BCEL 比其他三个方法具有大的覆盖概率, 其区间长度稍有差别. 因此, 我们的模拟说明 BCEL 效果最好. 此外, 当 n 增大时, 覆盖概率明显增大且区间长度缩短.

我们对正态 X 情形的模拟结果也适合于相应截断正态 X 的情形. 条件 (C.3.10.1) 和 (C3.10.2) 满足后者情形, 并且有趣的是该结果对正态 X 情形仍有效.

例3.12.2 考虑二次模型

$$Y=30(\beta^{\mathrm{T}}X-0.8)^2+\varepsilon,$$

其中 $\beta=(\sqrt{2}/2,\sqrt{3}/3,\sqrt{6}/6)^{\mathrm{T}}$, $\varepsilon\sim N(0,0.6^2)$, X 是三维随机变量, 各分量相互独立且都服从均匀分布 $U(0,1)$. 模拟结果展示在图 3.12.1 和表 3.12.2 中.

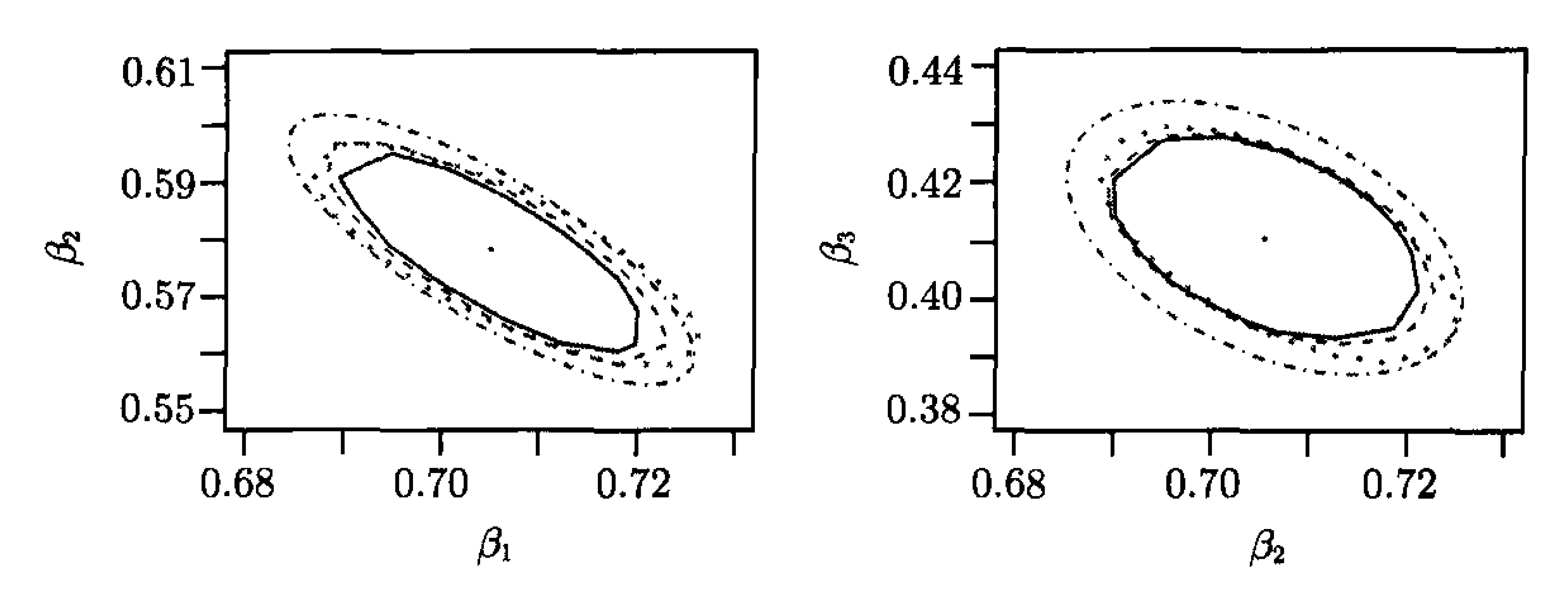


图 3.12.1　基于 BCEL(实线)、AEL(虚线)、EEL(点线) 和 NA(点–虚线) 及样本量 $n = 200$, 参数 (β_1, β_2) 的近似 0.95 置信域

表 3.12.2　参数 (β_1, β_2) 和 (β_1, β_3) 的近似 0.95 置信域的覆盖概率

n	(β_1, β_2)				(β_1, β_3)			
	BCEL	AEL	EEL	NA	BCEL	AEL	EEL	NA
150	0.910	0.908	0.910	0.904	0.910	0.908	0.910	0.906
200	0.922	0.914	0.920	0.906	0.922	0.914	0.920	0.908
300	0.948	0.944	0.928	0.922	0.948	0.944	0.928	0.926

从图 3.12.1 看出, BCEL 又一次显示其效果最好, 它的置信域比其他三个方法的置信域小. AEL 的效果也好. 三种经验似然置信域的覆盖概率有相当小的差别, 但它们都比 NA 实施效果好.

例3.12.3　考虑四维情形. 单指标模型为

$$Y = 15\sin(0.5\pi\beta^{\mathrm{T}}X) + \varepsilon,$$

其中 $\beta = (0.5, 0.5, 0.5, 0.5)^{\mathrm{T}}$, X 是四维随机变量, 各分量相互独立且都服从均匀分布 $U(0,1)$, ε 服从正态分布 $N(0, 0.4^2)$.

模拟结果放在表 3.12.3 中.

表 3.12.3　参数 $(\beta_1, \beta_2, \beta_3)$ 和 $(\beta_2, \beta_3, \beta_4)$ 的近似 0.95 置信域的覆盖概率

n	$(\beta_1, \beta_2, \beta_3)$				$(\beta_2, \beta_3, \beta_4)$			
	BCEL	AEL	EEL	NA	BCEL	AEL	EEL	NA
150	0.928	0.928	0.926	0.910	0.928	0.928	0.926	0.910
200	0.932	0.932	0.930	0.914	0.932	0.932	0.930	0.914
300	0.948	0.944	0.938	0.928	0.948	0.944	0.938	0.928

从表 3.12.3 的结果看出基于 NA 的覆盖概率明显比置信度 0.95 小. 在其他方法的比较中, BCEL 比 AEL 和 EEL 实施效果好.

根据上述模拟结果, 我们一致推荐 BCEL 方法, 因为它比所有其他方法的执行效果都好.

3.12.5 定理的证明

我们仅给出定理 3.12.1 和定理 3.12.2 的证明, 类似地, 可以证明定理 3.12.3. 首先需要给出下列引理.

引理 3.12.1 在定理 3.12.1 的条件下, 当 β_0 的第 r 个分量为正时, 有

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)}) \xrightarrow{D} N\big(0, V(\beta_0^{(r)})\big),$$

其中 $V(\beta_0^{(r)})$ 在条件 (C3.12.2) 中定义.

证 简单计算可得

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}(\beta_0^{(r)}) = V_n(\beta_0^{(r)}) + M_1 + M_2 + M_3, \tag{3.12.7}$$

其中

$$\begin{aligned}
V_n(\beta_0^{(r)}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i g'(\beta_0^{\mathrm{T}}X_i)J_{\beta_0^{(r)}}^{\mathrm{T}}\big[X_i - E(X_i|\beta_0^{\mathrm{T}}X_i)\big],\\
M_1 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\big[\hat{g}'(\beta_0^{\mathrm{T}}X_i;\beta_0) - g'(\beta_0^{\mathrm{T}}X_i)\big]J_{\beta_0^{(r)}}^{\mathrm{T}}X_i,\\
M_2 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}g'(\beta_0^{\mathrm{T}}X_i)J_{\beta_0^{(r)}}^{\mathrm{T}}\big\{\big[g(\beta_0^{\mathrm{T}}X_i) - \hat{g}(\beta_0^{\mathrm{T}}X_i;\beta_0)\big]X_i + \varepsilon_i E(X_i|\beta_0^{\mathrm{T}}X_i)\big\}\\
M_3 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[g(\beta_0^{\mathrm{T}}X_i) - \hat{g}(\beta_0^{\mathrm{T}}X_i;\beta_0)\big]\big[\hat{g}'(\beta_0^{\mathrm{T}}X_i;\beta_0) - g'(\beta_0^{\mathrm{T}}X_i)\big]J_{\beta_0^{(r)}}^{\mathrm{T}}X_i
\end{aligned}$$

由独立同分布的随机变量之和的中心极限定理可得

$$V_n(\beta_0^{(r)}) \xrightarrow{D} N(0, V(\beta_0^{(r)})). \tag{3.12.8}$$

类似于引理 3.10.4 的证明可以证得

$$M_l \xrightarrow{P} 0, \quad l = 1, 2, 3. \tag{3.12.9}$$

结合 (3.12.7)~(3.12.9) 式即可证得引理 3.12.1.

引理 3.12.2 在定理 3.12.1 的条件下, 当 β_0 的第 r 个分量为正时, 有

$$\frac{1}{n}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\hat{Z}_i(\beta_0^{(r)})^{\mathrm{T}} \xrightarrow{P} V_0(\beta_0^{(r)}),$$

其中 $V_0(\beta_0^{(r)})$ 在条件 (C3.12.2) 中定义.

证　记

$$R_{ni} = \left[\hat{g}'(\beta_0^{\mathrm{T}} X_i; \beta_0) - g'(\beta_0^{\mathrm{T}} X_i)\right] J_{\beta_0^{(r)}}^{\mathrm{T}} X_i \varepsilon_i + \left[g(\beta_0^{\mathrm{T}} X_i; \beta_0) - \hat{g}(\beta_0^{\mathrm{T}} X_i)\right] g'(\beta_0^{\mathrm{T}} X_i) J_{\beta_0^{(r)}}^{\mathrm{T}} X_i + \left[g(\beta_0^{\mathrm{T}} X_i; \beta_0) - \hat{g}(\beta_0^{\mathrm{T}} X_i)\right]\left[\hat{g}'(\beta_0^{\mathrm{T}} X_i; \beta_0) - g'(\beta_0^{\mathrm{T}} X_i)\right] J_{\beta_0^{(r)}}^{\mathrm{T}} X_i.$$

那么

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \hat{Z}_i(\beta_0^{(r)}) \hat{Z}_i^{\mathrm{T}}(\beta_0^{(r)}) \\
&= \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 g'(\beta_0^{\mathrm{T}} X_i)^2 J_{\beta_0^{(r)}}^{\mathrm{T}} X_i X_i^{\mathrm{T}} J_{\beta_0^{(r)}} + \frac{1}{n}\sum_{i=1}^{n} R_{ni} R_{ni}^{\mathrm{T}} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i g'(\beta_0^{\mathrm{T}} X_i) J_{\beta_0^{(r)}}^{\mathrm{T}} X_i R_{ni}^{\mathrm{T}} + \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i g'(\beta_0^{\mathrm{T}} X_i) R_{ni} X_i^{\mathrm{T}} J_{\beta_0^{(r)}} \\
&\equiv A_1 + A_2 + A_3 + A_4. \qquad (3.12.10)
\end{aligned}$$

由大数定律可得 $A_1 \xrightarrow{P} V_0(\beta_0^{(r)})$. 因此, 为证引理 3.12.2, 仅需证明 $A_l \xrightarrow{P} 0, l = 2, 3, 4$.

用 $A_{2,st}$ 表示 A_2 的 (s,t) 元素, $R_{ni,s}$ 表示 R_{ni} 的第 s 个分量. 那么由 Cauchy-Schwarz 不等式可得

$$|A_{2,st}| \leqslant \left(\frac{1}{n}\sum_{i=1}^{n} R_{ni,s}^2\right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^{n} R_{ni,s}^2\right)^{1/2} \qquad (3.12.11)$$

进一步计算可得

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} R_{ni,s}^2 &\leqslant cn^{-1} \sum_{i=1}^{n} \left[\hat{g}'(\beta_0^{\mathrm{T}} X_i; \beta_0) - g'(\beta_0^{\mathrm{T}} X_i)\right]^2 X_{is}^2 \varepsilon_i^2 \\
&\quad + cn^{-1} \sum_{i=1}^{n} \left[\hat{g}(\beta_0^{\mathrm{T}} X_i; \beta_0) - g(\beta_0^{\mathrm{T}} X_i)\right]^2 g'(\beta_0^{\mathrm{T}} X_i)^2 X_{is}^2 \\
&\quad + cn^{-1} \sum_{i=1}^{n} \left|\hat{g}(\beta_0^{\mathrm{T}} X_i; \beta_0) - g(\beta_0^{\mathrm{T}} X_i)\right|^2 \left|\hat{g}'(\beta_0^{\mathrm{T}} X_i; \beta_0) - g'(\beta_0^{\mathrm{T}} X_i)\right|^2 X_{is}^2 \\
&\equiv L_1 + L_2 + L_3.
\end{aligned}$$

设 $\hat{g}'(t;\beta_0)^{(-i)} = \hat{g}'(t;\beta_0) - \widetilde{W}_{ni}(t;\beta_0) Y_i$. 利用引理 3.10.2 和引理 3.10.3 可以推出, 对所有 $1 \leqslant i \leqslant n$, 有

$$E\left[\hat{g}'(\beta_0^{\mathrm{T}} X_i; \beta_0)^{(-i)} - g'(\beta_0^{\mathrm{T}} X_i)\right]^2 \leqslant ch_1^2 + c(nh_1^3)^{-1}.$$

用 $\hat{g}'(\beta_0^{\mathrm{T}}X_i;\beta_0)^{(-i)}+\widetilde{W}_{ni}(\beta_0^{\mathrm{T}}X_i;\beta_0)Y_i$ 代替 $\hat{g}'(\beta_0^{\mathrm{T}}X_i;\beta_0)$, 可得

$$\begin{aligned}E(L_1)&\leqslant cn^{-1}\sum_{i=1}^{n}E\big\{\big[\hat{g}'(\beta_0^{\mathrm{T}}X_i;\beta_0)^{(-i)}-g'(\beta_0^{\mathrm{T}}X_i)\big]^2X_{is}^2\varepsilon_i^2\big\}\\&\quad+cn^{-1}\sum_{i=1}^{n}E\big[\widetilde{W}_{ni}^2(\beta_0^{\mathrm{T}}X_i;\beta_0)Y_i^2X_{is}^2\varepsilon_i^2\big]\\&\leqslant ch_1^2+c(nh_1^3)^{-1}\to 0.\end{aligned}\tag{3.12.12}$$

因此, $L_1\xrightarrow{P}0$. 利用引理 3.10.2 和引理 3.10.3 可以证得 $L_2\xrightarrow{P}0$, $L_3\xrightarrow{P}0$. 这就这就证明了

$$\frac{1}{n}\sum_{i=1}^{n}R_{ni,s}^2\xrightarrow{P}0.$$

结合 (3.12.11) 式可知 $A_2\xrightarrow{P}0$. 类似地, 可以证明 $A_3\xrightarrow{P}0$, $A_4\xrightarrow{P}0$. 结合 (3.12.10) 式和 $A_1\xrightarrow{P}V_0(\beta_0^{(r)})$ 即完成了引理 3.12.2 的证明.

引理 3.12.3 在定理 3.12.1 的条件下, 当 β_0 的第 r 个分量为正时, 有

$$\max_{1\leqslant i\leqslant n}||\hat{Z}_i(\beta_0^{(r)})||=o_P\big(n^{1/2}\big).$$

证 用 $\hat{Z}_{i,s}(\beta_0^{(r)})$ 表示 $\hat{Z}_i(\beta_0^{(r)})$ 的第 s 个分量. 简单计算可得

$$\begin{aligned}\max_{1\leqslant i\leqslant n}|\hat{Z}_{i,s}(\beta_0^{(r)})|&\leqslant c\max_{1\leqslant i\leqslant n}|g'(\beta_0^{\mathrm{T}}X_i)X_{is}\varepsilon_i|\\&\quad+c\max_{1\leqslant i\leqslant n}|[\hat{g}'(\beta_0^{\mathrm{T}}X_i;\beta_0)-g'(\beta_0^{\mathrm{T}}X_i)]X_{is}\varepsilon_i|\\&\quad+c\max_{1\leqslant i\leqslant n}|[\hat{g}(\beta_0^{\mathrm{T}}X_i;\beta_0)-g(\beta_0^{\mathrm{T}}X_i)]g'(\beta_0^{\mathrm{T}}X_i)X_{is}|\\&\quad+c\max_{1\leqslant i\leqslant n}|[\hat{g}(\beta_0^{\mathrm{T}}X_i;\beta_0)-g(\beta_0^{\mathrm{T}}X_i)][\hat{g}'(\beta_0^{\mathrm{T}}X_i;\beta_0)-g'(\beta_0^{\mathrm{T}}X_i)]X_{is}|\\&\equiv B_1+B_2+B_3+B_4.\end{aligned}$$

由 Owen(1990) 给出的结果可知: 当 $E(Y^2)<\infty$ 时, 有 $\max_{1\leqslant i\leqslant n}|Y_i|=o\big(n^{1/2}\big)$, a.s. 因此可得 $B_1=o_P\big(n^{1/2}\big)$. 由 Markov 不等式和 (3.12.12) 式, 对任何 $\eta>0$, 有

$$\begin{aligned}P\big\{n^{1/2}B_2>\eta\big\}&\leqslant\sum_{i=1}^{n}P\big\{|[\hat{g}'(\beta_0^{\mathrm{T}}X_i;\beta_0)-g'(\beta_0^{\mathrm{T}}X_i)]X_{is}\varepsilon_i|>\eta\sqrt{n}\,\big\}\\&\leqslant\frac{1}{n\eta^2}\sum_{i=1}^{n}E\big\{\big[\hat{g}'(\beta_0^{\mathrm{T}}X_i;\beta_0)-g'(\beta_0^{\mathrm{T}}X_i)\big]^2X_{is}^2\varepsilon_i^2\big\}\\&\leqslant ch_1^2+c(nh_1^3)^{-1}\to 0.\end{aligned}$$

这就有 $B_2 = o_P(n^{1/2})$. 类似地, 可以证明: $B_3 = o_P(n^{1/2})$, $B_4 = o_P(n^{1/2})$. 这就完成了引理 3.12.3 的证明.

引理 3.12.4　在引理 3.12.1 的条件下, 有 $\lambda = O_P(n^{-1/2})$.

证　由引理 3.12.1 可得 $n^{-1}\sum\limits_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)}) = O_P(n^{-1/2})$. 结合引理 3.12.1, 并利用 Owen(1990) 中 (2.14) 式的证明方法即可证得该引理.

定理 3.12.1 的证　对 (3.12.1) 式应用 Taylor 公式并使用引理 3.12.2~ 引理 3.12.4, 可得

$$\hat{l}(\beta_0^{(r)}) = 2\sum_{i=1}^{n}\left\{\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0^{(r)}) - \frac{1}{2}\left[\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0^{(r)})\right]^2\right\} + o_P(1), \tag{3.12.13}$$

由 (3.12.2) 式可以推出

$$\begin{aligned}0 &= \sum_{i=1}^{n}\frac{\hat{Z}_i(\beta_0^{(r)})}{1+\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0^{(r)})}\\ &= \sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)}) - \sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\hat{Z}_i^{\mathrm{T}}(\beta_0^{(r)})\lambda + \sum_{i=1}^{n}\frac{\hat{Z}_i(\beta_0^{(r)})\left[\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0^{(r)})\right]^2}{1+\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0^{(r)})}.\end{aligned}$$

再应用引理 3.12.2~引理 3.12.4 可得

$$\sum_{i=1}^{n}[\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0^{(r)})]^2 = \sum_{i=1}^{n}\lambda^{\mathrm{T}}\hat{Z}_i(\beta_0^{(r)}) + o_P(1) \tag{3.12.14}$$

$$\lambda = \left(\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\hat{Z}_i^{\mathrm{T}}(\beta_0^{(r)})\right)^{-1}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)}) + o_P(n^{-1/2}). \tag{3.12.15}$$

结合 (3.12.13) 式可证得

$$\hat{l}(\beta_0^{(r)}) = \left(\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\right)^{\mathrm{T}}\left(\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\hat{Z}_i^{\mathrm{T}}(\beta_0^{(r)})\right)^{-1}\left(\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\right) + o_P(1).$$

由引理 3.12.2 可得

$$\hat{l}(\beta_0^{(r)}) = \left(V^{-\frac{1}{2}}(\beta_0^{(r)})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\right)^{\mathrm{T}}D_1(\beta_0^{(r)})\left(V^{-\frac{1}{2}}(\beta_0^{(r)})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\right) + o_P(1),$$

其中 $D_1(\beta_0^{(r)}) = V^{1/2}(\beta_0^{(r)})V_0^{-1}(\beta_0^{(r)})V^{1/2}(\beta_0^{(r)})$.

设 $\widetilde{D} = \mathrm{diag}(w_1, \cdots, w_{p-1})$, 其中 $w_i(1 \leqslant i \leqslant p-1)$ 是 $V_0^{-1}(\beta_0^{(r)})V(\beta_0^{(r)})$ 的特征根. 注意到 $D_1(\beta_0^{(r)})$ 与 $V_0^{-1}(\beta_0^{(r)})V(\beta_0^{(r)})$ 有同样的特征根. 那么正交矩阵 Q, 使得 $Q^{\mathrm{T}}\widetilde{D}Q = D_1(\beta_0^{(r)})$. 这里

$$\hat{l}(\beta_0^{(r)}) = \left(QV^{-\frac{1}{2}}(\beta_0^{(r)})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\right)^{\mathrm{T}}\widetilde{D}\left(QV^{-\frac{1}{2}}(\beta_0^{(r)})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\right) + o_P(1).$$

由引理 3.12.1 可得

$$QV^{-\frac{1}{2}}(\beta_0^{(r)})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)}) \xrightarrow{\mathcal{L}} N(0, I_{p-1}), \tag{3.12.16}$$

其中 I_{p-1} 为 $(p-1)\times(p-1)$ 单位矩阵. 结合上述两个结果即可证得定理 3.12.1.

定理 3.12.2 的证 回顾 (3.12.4) 式中的 $\hat{l}_{\mathrm{AEL}}(\beta_0^{(r)})$, 可以推出

$$\hat{l}_{\mathrm{AEL}}(\beta_0^{(r)}) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\right)^{\mathrm{T}}\hat{V}^{-1}(\beta_0^{(r)})\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\right) + o_P(1).$$

类似于引理 3.12.2 的证明方法, 可以证明 $\hat{V}(\beta_0^{(r)}) \xrightarrow{P} V(\beta_0^{(r)})$. 该式与 (3.12.16) 式结合即证明了 $\hat{l}_{\mathrm{AEL}}(\beta_0^{(r)})$ 渐近服从自由度为 $p-1$ 的 χ^2 分布.

3.13 拟合优度检验

Xia 等 (2004) 研究了模型 (3.1.1) 的拟合优度检验问题, 构造了 Cramér-von Mises 检验统计量, 并用 bootstrap 方法提供 p 值和处理非参数估计的偏差问题. 本节取材主要来源于 Xia 等 (2004).

3.13.1 Cramér-von Mises 检验

本节使用 u 表示标量, 使用 x 表示向量. 考虑下列检验:

H_0 : 存在常数向量 β_0 和联系函数 $g(\cdot)$, 使得 $E[Y-g(\beta_0)|X]=0$ 几乎处处成立.

H_0 成立当且仅当

$$E[Y-g(\beta_0^{\mathrm{T}}X)|X]=0, \quad \text{a.s.} \tag{3.13.1}$$

为构造 Cramér-von Mises 统计量, 进一步注意到 (3.13.1) 式等价于

$$E\{[Y-g(\beta_0^{\mathrm{T}}X)]I(X<x)\}\equiv 0, \tag{3.13.2}$$

其中 $X<x$ 表示 X 的每一个分量小于 x 的相应分量. 假设 $\{(X_i,Y_i), 1\leqslant i\leqslant n\}$ 是来自模型 (3.1.1) 的 iid 样本, 并设诸 $\hat{Y}_i$ 是用某种非参数方法得到的拟合值. 相应于 (3.13.2) 式, 定义残差符号经验过程

$$R_n(x)=n^{-1/2}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)I(X_i<x).$$

为计算拟合值 $\hat{Y}_i$, 需要估计 $g(\cdot)$ 和 β_0. 对固定的 β_0, 我们使用 $g(\cdot)$ 的局部线性估计 $\hat{g}(\cdot;\beta_0)$, 见 (3.10.4) 式. 此外, 前面介绍的方法都可以用来估计 β_0, 这里使

用 β_0 的估计方程估计 $\hat{\beta}$(见 3.10 节). 有了 $g(\cdot)$ 和 β_0 的估计之后, 就可以得到 Y_i 的拟合值 $\hat{Y}_i = \hat{g}(\hat{\beta}^{\mathrm{T}} X_i; \hat{\beta})$, 从而得到 $R_n(x)$ 的值.

为避免 $\hat{g}(\hat{\beta}^{\mathrm{T}} X_i; \hat{\beta})$ 的分母的值接近于 0 而引起的麻烦, 一个常用的方法是去掉边界点, 即去掉满足 $||X_i|| > c$ 的点, 其中 c 为某个正的常数. 如 (Härdle et al., 1993; Weisberg and Welsh 1994). 这里仅删除 $\hat{\beta}^{\mathrm{T}} X_i \in \mathcal{D}$ 的观察值, 其中 $\mathcal{D}$ 是 $\beta_0^{\mathrm{T}} X$ 的具有正密度的紧集. 进一步定义残差符号经验过程

$$R_{\mathcal{D}}(x) = n^{-1/2} \sum_{\hat{\beta}^{\mathrm{T}} X_i \in \mathcal{D}} (Y_i - \hat{Y}_i) I(X_i < x). \tag{3.13.3}$$

相应地, 定义 $I_{\mathcal{D}}(X_i < x) = I(\beta_0^{\mathrm{T}} X \in \mathcal{D}) I(X_i < x)$. 按照 Xia 等 (2004) 的讨论, 设

$$\begin{aligned} H(x) = \Big\{ & I_{\mathcal{D}}(X < x) - E\left(\{g'(\beta_0^{\mathrm{T}} X) I_{\mathcal{D}}(X < x)[X - E(X|\beta_0^{\mathrm{T}} X)]\}^{\mathrm{T}}\right) l(X, g, \beta_0) \\ & - E[I_{\mathcal{D}}(X < x)|\beta_0^{\mathrm{T}} X] \Big\} \varepsilon, \end{aligned}$$

$$l(x, g, \beta_0) = \left[\int w(z)[z - \mu(\beta_0^{\mathrm{T}} z)]^{\otimes 2} g'(\beta_0^{\mathrm{T}} z)^2 f_X(z) \mathrm{d}z \right]^{-} [x - \mu(\beta_0^{\mathrm{T}} x)],$$

其中 $\mu(\beta_0^{\mathrm{T}} x) = E(X|\beta_0^{\mathrm{T}} X = \beta_0^{\mathrm{T}} x)$, $f_X(x)$ 为 X 的密度函数.

记 $\bar{\mathfrak{R}}^{\otimes p} = [-\infty, \infty]^{\otimes p}$, $D(\bar{\mathfrak{R}}^{\otimes p})$ 是 Skorokhod 空间. 用 “$\Rightarrow$” 表示弱收敛. 在适当的条件下, Xia 等 (2004) 证明了下列结果: 当 H_0 为真时, 在 $D(\bar{\mathfrak{R}}^{\otimes p})$ 上, 有

$$R_{\mathcal{D}}(x) + B_{\mathcal{D}}(x) \Rightarrow Q(x), \tag{3.13.4}$$

其中 $B_{\mathcal{D}}(x) = n^{1/10} E[g''(\beta_0^{\mathrm{T}} X) I_{\mathcal{D}}(X < x)]/2$, $Q(x)$ 是均值为 0 的 Gauss 过程, 其协方差函数为 $E[Q(x_1)Q(x_2)] = E[H(x_1)H(x_2)]$.

残差符号经验过程 $R_{\mathcal{D}}(x)$ 有一个偏差项 $B_{\mathcal{D}}(x)$. 如果我们为了检验而想使用过程 $R_{\mathcal{D}}(x)$, 那么我们必须去掉这个偏差项. 因此, 纠偏的统计量定义为

$$\mathrm{CCV}_{\mathcal{D}} = \int_{-\infty}^{\infty} [R_{\mathcal{D}}(x) + B_{\mathcal{D}}(x)]^2 F_n(\mathrm{d}x),$$

其中 $F_n(x)$ 为 X 的经验分布函数. 由 (3.13.4) 式可得

$$\mathrm{CCV}_{\mathcal{D}} \xrightarrow{D} \int_{-\infty}^{\infty} Q^2(x) F(\mathrm{d}x).$$

其中 $F(x)$ 为 X 的分布函数.

正如前面已经解释的, 偏差项在实际中会引起麻烦. 原则上, 可以使用通常的方法估计它, 然而, 需要选择另一个带宽, 这有一定难度. 此外, 渐近分布仍然依赖于未知函数的衍生项, 要估计它也有困难. 给出分布的近似形式也很难做到. 鉴于此, Xia 等 (2004) 利用 bootstrap 方法得到偏差的一个估计, 并仿真未知分布.

3.13.2 bootstrap 方法

本节利用回归模型的广义 bootstrap 方法进行推断, 见文献 (Härdle and Mammen, 1993). 假设 $\{(X_i, Y_i), 1 \leqslant i \leqslant n\}$ 是来自模型 (3.1.1) 的 iid 样本. 在 H_0 下, 首先用 3.13.1 小节的方法估计 β_0 和 $g(\cdot)$, 然后从模型

$$Y_i^* = \hat{g}(\hat{\beta}^{\mathrm{T}} X_i; \hat{\beta}) + \varepsilon_i^*, \quad i = 1, \cdots, n \tag{3.13.5}$$

产生独立 bootstrap 观察值, 其中 $\varepsilon_i^* = (Y_i - Y_i^*)\epsilon_i^*$, 诸 ϵ_i^* 是 iid 随机变量, 每一个都具有均值 0, 方差 1, 所有阶矩, 并独立于 $\{(X_i, Y_i), 1 \leqslant i \leqslant n\}$. 可以用 3.13.1 小节的方法重新估计 $\hat{\beta}$ 和 $\hat{g}(u; \hat{\beta})$, 并用 $\hat{\beta}^*$ 和 $\hat{g}^*(u; \hat{\beta}^*)$ 分别记相应的估计. $R_{\mathcal{D}}(x)$ 的 bootstrap 形式是

$$R_{\mathcal{D}}^*(x) = n^{-1/2} \sum_{\hat{\beta}^{\mathrm{T}} X_i \in \mathcal{D}} (Y_i^* - \hat{Y}_i)^* I(X_i < x). \tag{3.13.6}$$

1. 纠偏思想

下面给出纠偏思想的解释. 根据定理 3.10.4 的证明可得: 在一些假定下,

$$\hat{g}(u; \hat{\beta}) = g(u) + \frac{1}{2} h^2 \kappa_2 g''(u) + \frac{1}{nf(u)} \sum_{i=1}^{n} \varepsilon_i K_h(\beta_0^{\mathrm{T}} X_i - u) + o_P(h^2),$$

其中 $K_h(\cdot) = h^{-1} K(\cdot/h)$, $K(\cdot)$ 为核函数, $h = h_n$ 为收敛于 0 的带宽序列. 对每个 bootstrap 样本, 有

$$\begin{aligned} \hat{g}^*(u; \hat{\beta}) = g(u) + h^2 \kappa_2 g''(u) + \frac{1}{nf(u)} \sum_{i=1}^{n} \varepsilon_i K * K_h(\beta_0^{\mathrm{T}} X_i - u) \\ + \frac{1}{nf(u)} \sum_{i=1}^{n} \varepsilon_i^* K_h(\beta_0^{\mathrm{T}} X_i - u) + o_P(h^2), \end{aligned} \tag{3.13.7}$$

其中 $K * K$ 表示 K 的卷积. $\hat{g}^*(u; \hat{\beta})$ 的偏差为

$$\begin{aligned} & E\left[\hat{g}^*(u; \hat{\beta}) - \hat{g}(u; \hat{\beta}) \middle| (X_1, Y_1), \cdots, (X_n, Y_n)\right] \\ = & \frac{1}{2} h^2 \kappa_2 g''(u) + \frac{1}{nf(u)} \sum_{i=1}^{n} \varepsilon_i [K * K_h(\beta_0^{\mathrm{T}} X_i - u) - K_h(\beta_0^{\mathrm{T}} X_i - u)] + o_P(h^2), \end{aligned} \tag{3.13.8}$$

注意到上式右端第二项的阶为 $O_P(h^2)$, 因此 (3.13.8) 式蕴涵 $\hat{g}^*(u; \hat{\beta})$ 和 $\hat{g}(u; \hat{\beta})$ 有不同的偏差项. 因此, 如果我们试图作关于回归函数 $g(\cdot)$ 的逐点推断, 就必须使

用另一个带宽, 且超光滑回归函数使得 (3.13.8) 式右端第二项的阶为 $o_P(h^2)$. 见 Härdle(1990). 然而, 在我们的问题中以残差符号经验过程的和可以减少差别, 即

$$\sum_{i=1}^{n} \varepsilon_i [K * K_h(\beta_0^{\mathrm{T}} X_i - u) - K_h(\beta_0^{\mathrm{T}} X_i - u)] = o_P(n^{1/2}), \tag{3.13.9}$$

这是因为 $\int [K*K(u) - K(u)]\mathrm{d}u = 0$. 由 (3.13.9) 式可以证明 $R^*_{\mathcal{D}}(x)$ 与 $R_{\mathcal{D}}(x)$ 有相同的渐近偏差项. 注意到 (3.13.8) 式, 偏差 $E\left[\hat{g}^*(u;\hat{\beta}) - \hat{g}(u;\hat{\beta}) | (X_1, Y_1), \cdots, (X_n, Y_n)\right]$ 可以用样本平均得到. 因此, 容易计算和排除 $R^*_{\mathcal{D}}(x)$ 和 $R_{\mathcal{D}}(x)$ 中的偏差项.

2. 渐近分布

记 $B_i = E\left[\hat{g}^*(u;\hat{\beta}) - \hat{g}(u;\hat{\beta}) | (X_1, Y_1), \cdots, (X_n, Y_n)\right]$, $\widetilde{Y}_i = \hat{Y}_i - B_i$, $\widetilde{Y}_i^* = \hat{Y}_i^* - B_i$, 其中 $\hat{Y}_i^* = \hat{g}^*(\hat{\beta}^{*T} X_i; \hat{\beta}^*)$. 那么可以把 $\widetilde{Y}_i$ 和 $\widetilde{Y}_i^*$ 看成纠偏拟合值. 设

$$\widetilde{R}_{\mathcal{D}}(x) = n^{-1/2} \sum_{\beta^{\mathrm{T}} X_i \in \mathcal{D}} (Y_i - \widetilde{Y}_i) I(X_i < x).$$

它的 bootstrap 相似为

$$\widetilde{R}^*_{\mathcal{D}}(x) = n^{-1/2} \sum_{\hat{\beta}^{\mathrm{T}} X_i \in \mathcal{D}} (Y_i^* - \widetilde{Y}_i^*) I(X_i < x).$$

Xia 等 (2004) 在适当的条件下证明了下列结论: 当 H_0 为真时, 在 $D(\Re^{\otimes p})$ 中, 有

$$\widetilde{R}_{\mathcal{D}}(x) \Rightarrow Q(x), \quad \widetilde{R}^*_{\mathcal{D}}(x) \Rightarrow Q(x), \tag{3.13.10}$$

其中 $Q(x)$ 在 (3.13.4) 式中定义.

因为 $\widetilde{R}^*_{\mathcal{D}}(x)$ 与 $\widetilde{R}_{\mathcal{D}}(x)$ 有相同的渐近分布, 所以可以使用它们来检验假设 H_0. 考虑下列 Cramér-von Mises 统计量

$$\mathrm{CVS}_{\mathcal{D}} = \int_{-\infty}^{\infty} \widetilde{R}^2_{\mathcal{D}}(x) F_n(\mathrm{d}x), \quad \mathrm{CVS}^*_{\mathcal{D}} = \int_{-\infty}^{\infty} \widetilde{R}^{*2}_{\mathcal{D}}(x) F_n(\mathrm{d}x).$$

由 (3.13.10) 知, $\mathrm{CVS}_{\mathcal{D}}$ 与 $\mathrm{CVS}^*_{\mathcal{D}}$ 有相同的渐近分布. 因此, 可以用 $\mathrm{CVS}^*_{\mathcal{D}}$ 模仿 $\mathrm{CVS}_{\mathcal{D}}$ 的分布.

3. 带偏差项的检验统计量

注意到在 $D(\Re^{\otimes p})$ 中有 $R^*_{\mathcal{D}}(x) + B_{\mathcal{D}}(x) \Rightarrow Q(x)$, 再由 (3.13.4) 式知 $R^*_{\mathcal{D}}$ 与 $R_{\mathcal{D}}(x)$ 有相同的偏差项, 且去掉偏差项后也有相同的渐近分布, 为简化计算, 可以构造检验统计量 (带偏差项)

$$\mathrm{CVT}_{\mathcal{D}} = \int_{-\infty}^{\infty} R^2_{\mathcal{D}}(x) F_n(\mathrm{d}x), \quad \mathrm{CVT}^*_{\mathcal{D}} = \int_{-\infty}^{\infty} R^{*2}_{\mathcal{D}}(x) F_n(\mathrm{d}x).$$

基于 $\mathrm{CVS}_{\mathcal{D}}$ 和 $\mathrm{CVS}^*_{\mathcal{D}}$, 可以对 H_0 进行假设检验.

4. 单指标检验统计量

记 $r(x)=E[Y-g(\beta_0^{\mathrm{T}}X)|X=x]$. 通常可以由 $E[r(X)|\beta^{\mathrm{T}}X]$ 判断是否 $r(X)\neq 0$. 记 $r_\beta(u)=E[Y-g(\beta_0^{\mathrm{T}}X)|\beta^{\mathrm{T}}X=u]$, $\beta_1=\arg\min_{\beta:||\beta||=1}E[Y-g(\beta_0^{\mathrm{T}}X)-r_\beta(\beta^{\mathrm{T}}X)]^2$. 那么, $r_{\beta_1}(u)$ 恰好是投影寻踪回归的第二个分量. 可以重复 β_0 的估计方法来估计 β_1, 其估计量记为 $\hat{\beta}_1$. 类似地, 我们可以估计 bootstrap 相似, 记为 $\hat{\beta}_1^*$. 设

$$\breve{R}_{\mathcal{D}}(u)=n^{-1/2}\sum_{\hat{\beta}^{\mathrm{T}}X_i\in\mathcal{D}}(Y_i-\widetilde{Y}_i)I(\hat{\beta}_1^{\mathrm{T}}X_i<u),$$

$$\breve{R}_{\mathcal{D}}^*(u)=n^{-1/2}\sum_{\hat{\beta}^{\mathrm{T}}X_i\in\mathcal{D}}(Y_i^*-\widetilde{Y}_i^*)I(\hat{\beta}_1^{*\mathrm{T}}X_i<u).$$

类似地, 记

$$\mathrm{C\check{V}S}_{\mathcal{D}}=\int_{-\infty}^{\infty}\breve{R}_{\mathcal{D}}^2(u)F_{n,\hat{\beta}_1}(\mathrm{d}u),\quad \mathrm{C\check{V}S}_{\mathcal{D}}^*=\int_{-\infty}^{\infty}\breve{R}_{\mathcal{D}}^{*2}(u)F_{n,\hat{\beta}_1}(\mathrm{d}u),$$

其中 $F_{n,\beta}(u)=\dfrac{1}{n}\sum_{i=1}^{n}I(\beta^{\mathrm{T}}X_i<u)$. 基于 $\mathrm{C\check{V}S}_{\mathcal{D}}$ 和 $\mathrm{C\check{V}S}_{\mathcal{D}}^*$ 也可以对假设 H_0 进行检验.

3.13.3 得分检验

本节介绍 Stute 和 Zhu(2005) 提出的检验. 在零假设下所考虑的单指标模型具有形式 (3.1.1). 噪声变量 ε 满足

$$E(\varepsilon|X)=E(\varepsilon|\beta_0^{\mathrm{T}}X)=0,\tag{3.13.11}$$

这等价于满足

$$E(Y|X)=E(\varepsilon|\beta_0^{\mathrm{T}}X)=g(\beta_0^{\mathrm{T}}X).\tag{3.13.12}$$

注意到 (3.13.11) 式允许 ε 依赖于 X, 以致 (3.1.1) 式包含异方差误差. (3.13.12) 的第一式描述了单指标模型的投影寻踪特征.

为在一定的技术水平上引出检验方法, 假定 $\beta_0^{\mathrm{T}}X$ 具有连续分布函数 F^{β_0}, 即

$$F(u)\equiv F^{\beta_0}(u)=P(\beta_0^{\mathrm{T}}X\leqslant u),\quad u\in\mathbf{R}.$$

用 F^{-1} 记 F 的分位函数:

$$F^{-1}(v)=\inf\{u\in\mathbf{R}:F(u)\geqslant v\},\quad 0<v<1.$$

令 $V\equiv F(\beta_0^{\mathrm{T}}X)$. 由 F 的连续性知 V 在区间 $(0,1)$ 上服从均匀分布. 记

$$\psi=g\circ F^{-1},$$

(3.1.1) 式变为 (以概率 1)

$$Y = \psi(V) + \varepsilon.$$

用回归的术语, 这可以表示为

$$m(x) = E(Y|X) = g(\beta_0^{\mathrm{T}} X) = \psi(v),$$

其中

$$v = F(\beta_0^{\mathrm{T}} x), \quad \psi(v) = E[Y|F(\beta_0^{\mathrm{T}} X) = v].$$

因此, 在 $(0,1)$ 上 $\psi(v)$ 的核估计为

$$\psi_n(u) = \frac{1}{n}\sum_{i=1}^{n} Y_i K_h(V_i - v),$$

其中 $K_h(\cdot) = h^{-1}K(\cdot/h)$, K 为积分非 1 的对称核, $h = h_n > 0$ 为带宽. 随机变量 $V_i = F(\beta_0^{\mathrm{T}} X)$ 是来自 $(0,1)$ 上均匀分布的随机变量. 因为 F 和 β_0 都是未知的, ψ_n 不能成为最后的估计. 为此, 我们用某个估计 $\hat{\beta}$ 代替 β_0, 并用 $\{\hat{\beta}^{\mathrm{T}} X_i, 1 \leqslant i \leqslant n\}$ 的经验分布函数 F_n 代替 F. 得到

$$\hat{V}_i = F_n(\hat{\beta}^{\mathrm{T}} X_i), \quad 1 \leqslant i \leqslant n,$$

以及相应的估计量

$$\hat{\psi}(v) = \frac{1}{n}\sum_{i=1}^{n} Y_i K_h(\hat{V}_i - v),$$

它被称为对称的最紧邻估计. 诸 $\hat{V}_i$ 是与投影值 $\hat{\beta}^{\mathrm{T}} X$ 有关的标准秩. 因为这些值依赖于随机的 $\hat{\beta}$, 关于秩统计量的现有结果不能容易地用来分析我们的检验统计量. 尤其诸 $\hat{V}_i$ 出现在光滑函数 $\hat{\psi}$ 的部分. 比这更糟糕的是, 我们必须在每一 $\hat{V}_j$ 处估计 $\hat{\psi}$, 最后导致残差

$$\hat{\varepsilon}_j = Y_j - \hat{\psi}(\hat{V}_j), \quad 1 \leqslant j \leqslant n.$$

实际上, 为减少偏差, 我们将考虑与 $\hat{\psi}$ 同样方法计算得到的 $\hat{\psi}^{(j)}$, 但其中删除了第 j 个观察值. 因此, 定义残差

$$\hat{\varepsilon}_j = Y_j - \hat{\psi}^{(j)}(\hat{V}_j), \quad 1 \leqslant j \leqslant n.$$

最后的检验统计量将具有形式

$$\hat{T} = n^{-1/2}\sum_{j=1}^{n} \hat{\varepsilon}_j W_j.$$

权 W_j 具有形式 $W_j = W(X_j)$, 其中 $W(\cdot)$ 是定义在 $\mathbf{R}^p$ 上的光滑函数. 后面再讨论如何选取 $W(\cdot)$. 在零模型 (3.13.11) 中, 我们期望 $\hat{T}$ 具有

$$T_n = n^{-1/2}\sum_{j=1}^{n}\varepsilon_j W_j$$

的性质. 因为 ε_j 与 W_j 正交, 所以 T_n 是中心化的变量. 因此, 我们也希望在 (3.13.11) 下 $\hat{T}$ 在 0 附近波动. 在 (局部) 备择下, ε_j 也包含不与诸 W_j 正交的量. 如果以适当的方式选择 $W(\cdot)$, 这将保证检验的非平凡功效.

更具体地, 我们将首先考虑模型

$$Y_{ni} = g(\beta_0^{\mathrm{T}}X_i) + n^{-1/2}s(X_j) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{3.13.13}$$

其中 $\{(X_i, \varepsilon_i), 1 \leqslant i \leqslant n\}$ iid, 并满足

$$E(\varepsilon_i|X_i) = 0, \quad i = 1, \cdots, n, \tag{3.13.14}$$

这里函数 $g(\cdot)$ 和参数 β_0 仍未指定, 函数 $s(\cdot)$ 可以指定, 也可以不指定. 当 $s(\cdot) \equiv 0$ 时, 单指标模型成立. 对具体的备择假设, 后面将讨论如何选择 $W(\cdot)$ 以得到最大的功效. 需要说明的是, 在实际应用中估计量 $\hat{\beta}$ 可以用前面给出的任一种方法得到, 没有具体要求.

Stute 和 Zhu(2005) 证明了下列结果: 在一些正则条件下, 有

$$\hat{T} = \mu + n^{-1/2}\sum_{j=1}^{n}\varepsilon_j[W_j - \bar{W}(V_j)] + o_P(1), \tag{3.13.15}$$

其中 $\bar{W}(v) = E[W(X)|V = v]$, $0 < v < 1$; 因此, 由中心极限定理可得

$$\hat{T} \xrightarrow{D} N(\mu, \sigma^2),$$

其中

$$\sigma^2 = E\{[\varepsilon^2[W(X) - \bar{W}(V)]^2\}.$$

$$\mu = E\{[s(X) - E(s(X)|V)]W(X)\},$$

这个漂移包含来之由 $\beta_0^{\mathrm{T}}X$ 张成的变量空间 $s(X)$ 的偏差. 在单指标模型下, 括号内等于 0, 且因此 μ 等于 0. 也有, 如果 $W(X)$ 不通过 $\beta_0^{\mathrm{T}}X$ 依赖于 X, 那么 μ 也等于 0. 渐近方差既不依赖于 $s(\cdot)$, 也不依赖于未知量 $g(\cdot)$. 可以得到 σ^2 的一个相合估计

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}\hat{\varepsilon}_j^2[W(X_j) - \hat{\bar{W}}^{(j)}(\hat{V}_j)]^2,$$

其中 $\hat{\bar{W}}^{(j)}$ 类似于 $\hat{\psi}^{(j)}$ 的定义; 仅在对称的最紧邻估计的定义中用 $W(X_i)$ 代替 Y_i 即可. 置

$$\hat{T}^* = \hat{T}/\hat{\sigma}^2,$$

那么, 可以得到

$$\hat{T}^* \xrightarrow{D} N(C, 1),$$

其中 $C = \mu/\sigma^2$. 如果

$$\hat{T}^* \geqslant z_{1-\alpha/2},$$

那么拒绝零模型, 其中 $z_{1-\alpha/2}$ 为标准正态分布函数 Φ 的 $1-\alpha/2$ 分位点. 因此对局部备择 (3.13.13), $\hat{T}^*$ 的渐近功效等于 $1-[\Phi(C+\lambda)-\Phi(C-\lambda)]$. 这是 $|C|$ 的单调函数. 因此, 我们可以选择 W 使得 C^2 尽可能大. 如果我们写

$$C^2 = C(s, W) = \frac{\mu^2(s, W)}{\sigma^2(W)},$$

当 ε 与 X 独立时, 容易确定所考虑问题的最优解. 即, 如果同方差情况成立, 那么上面的比等于

$$\frac{\mu^2(s, W)}{E(\varepsilon^2)E[W(X) - \bar{W}(V)]^2},$$

并且由 Cauchy-Schwarz 不等式立刻得到最优权函数 $W_0(\cdot) = s(\cdot)$.

3.14 变量选择

Kong 和 Xia(2007) 考虑了单指标模型的变量选择问题. 他们提出了一种新的变量选择方法, 称之为分离交叉验证. 进一步分析得到: 这个方法有更好的有限样本性质, 且比 leave-m-out 交叉验证更容易计算.

3.14.1 Leave-m-out 交叉验证

我们使用类似于 Xia(2007) 的记号. 用 $\mathcal{S}$ 表示 $\{1, \cdots, p\}$ 的所有非空子集. 对任何 $r \in \mathcal{S}$, 设 d_r 是 r 的势, 并设 β_r 和 X_r 为两个分别包含 β 和 X 的分量的 $d_r \times 1$ 列向量. 用 β 表示使 $E[Y - g_r(\beta^{\mathrm{T}} X_r)]^2$ 达到最小的向量. 用 $\mathcal{M}_r$ 表示相应的单指标模型

$$Y = g_r(\beta^{\mathrm{T}} X_r) + \varepsilon_r, \tag{3.14.1}$$

其中 $\varepsilon_r = Y - g_r(\beta^{\mathrm{T}} X_r)$.

如果我们知道真参数 β_0 的哪些分量是 0, 那么就可以把模型 $\mathcal{M}_r$ 分成两类. 在一类中, 至少一个协变量其在模型 (3.1.1) 中的非零系数在 X_r 缺失. 在另一类中,

X_r 包含所有非零系数的协变量. 第二类中具有协变量最小数 d_0 的那一类被定义为最优模型, 并记为 $\mathcal{M}_{r_0}$.

假设 $\{(X_i,Y_i),1\leqslant i\leqslant n\}$ 是来自模型 (3.1.1) 的 iid 样本. 考虑 $r\supset r_0$ 的模型 $\mathcal{M}_r$. 为保证估计的相合性, 本章假定: 对 $\beta_{0,r}$ 的邻域内的所有 β, 随机变量 $\beta^{\mathrm{T}}X_r$ 具有密度函数, 其中 $\beta_{0,r}$ 表示包含 β_0 的列向量. Horowitz 和 Härdle(1996) 给出了详细的讨论. 前面几节介绍的方法都可以用来估计 $\beta_{0,r}$, 且得到的估计量都具有 $\sqrt{n}$ 相合性. 下面我们介绍交叉验证变量选择方法.

在交叉验证方法中, 把数据分成两个集: 训练集 S^c 和检验集 S. 训练集用来估计所有候选模型和最优预测模型, 检验集是首选模型. 注意到在 (3.14.1) 中, 对任何 $r\supset r_0$ 有 $\beta=\beta_{0,r}$. 对这样的 r 和具有势 m 的 $s\in\{1,\cdots,n'\}$, 我们首先使用 $\{(X_i,Y_i),1\leqslant i\leqslant n,i\bar{\in}S\}$ 构造模型 (3.14.1) 中 $\beta_{0,r}$ 的最小平均方差估计 $\hat{\beta}_r^{\setminus s}$, 然后用局部线性光滑估计 $g_r(x|\hat{\beta}_r^{\setminus s})$, 其中 $g_r(x|\beta)=E(Y|\beta^{\mathrm{T}}X_r=\beta^{\mathrm{T}}x)$. 即, $g_r(u|\hat{\beta}_r^{\setminus s})$ 的局部线性估计定义为

$$\hat{g}_r^{\setminus s}(x|\hat{\beta}_r^{\setminus s})=\sum_{j\bar{\in}S}V_{j,r}(x|\hat{\beta}_r^{\setminus s})Y_j\Big/\sum_{i\bar{\in}S}V_{i,r}(x|\hat{\beta}_r^{\setminus s}),\tag{3.14.2}$$

其中

$$V_{j,r}(x|\beta)=K_h(\beta^{\mathrm{T}}(X_{j,r}-x))\left\{S_{r,2}^{\setminus s}(x|\beta)-[\beta^{\mathrm{T}}(X_{j,r}-x)/h]S_{r,1}^{\setminus s}\right\},$$

$$S_{r,k}^{\setminus s}(x|\beta)=\sum_{j\bar{\in}S}K_h(\beta^{\mathrm{T}}(X_{j,r}-x))[\beta^{\mathrm{T}}(X_{j,r}-x)/h]^k,\quad k=1,2.$$

定义 leave-m-out 交叉验证函数

$$\mathrm{CV}_r(m)=m^{-1}\binom{n'}{n}^{-1}\sum_S{}'\sum_{i\in S}[Y_i-\hat{g}_r^{\setminus s}(X_{i,r}|\hat{\beta}_r^{\setminus s})]^2,$$

其中 $\sum_S{}'$ 表示在所有势为 m 的可能子集 $S\subset\{1,\cdots,n'\}$ 上求和. 后面将使用 $\sum_{i,s}{}'$ 表示 $\sum_S{}'\sum_{i\in S}$. 最后所选择的模型是 $\mathrm{CV}_r(m)$ 具有最小值的模型 $\mathcal{M}_r$.

Kong 和 Xia(2007) 证明了下列结果: 在一些正则条件下, 如果 $m\to\infty$ 及 $m/n\to c\in[0,1)$, 且 $h\propto n^{-1/5}$, 则对任何 $r\supset r_0$ 和 $\delta_d\equiv d_r-d_0$, 有

$$\lim_{n\to\infty}P\left\{\mathrm{CV}_r(m)>\mathrm{CV}_{r_0}(m)\right\}=P\left\{\chi^2(\delta_d)>\frac{(2-3c)\delta_d}{1-c}\right\}.\tag{3.14.3}$$

由 (3.14.3) 知, leave-m-out 交叉验证是相合的, 即 $\lim\limits_{n\to\infty}P\{\mathrm{CV}_r(m)>\mathrm{CV}_{r_0}(m)\}=1$, 必须 $c\in[2/3,1)$. 虽然在情况 $c=1$ 下没有结论, 但推测相合性不成立, 因当 $n_c\equiv n-m=o(n)$ 时, 即训练集的样本非常小于 n 时, $\hat{\beta}_r^{\setminus s}$ 的 $\sqrt{n}$ 相合性不再成立.

因为线性模型中的参数可用小样本很好地估计, 所以 leave-m-out 交叉验证分离数据对线性模型是可接受的. 然而, 用作 leave-m-out 交叉验证的训练集的样本量对非参数光滑方法通常太小. leave-m-out 交叉验证的另一个缺点是它负担繁重的计算量, 因为需要计算组合数 $\begin{pmatrix} n' \\ m \end{pmatrix}$. 为减少重负, Monte-Carlo leave-m-out 交叉验证应运而生, 即通过极小化

$$\mathrm{CV}_r^{\mathrm{mc}}(m) = \sum_{s\in\mathcal{R}}\sum_{i\in S}[Y_i - \hat{g}_r^{\backslash s}(X_{i,r}|\hat{\beta}_r^{\backslash s})]^2, \tag{3.14.4}$$

来选择一个模型, 其中 $\mathcal{R}$ 是 $\{1,\cdots,n'\}$ 的容量为 m 的子集. 在线性回归模型中, 已经证明这个方法的性能类似于 leave-m-out 交叉验证; 见 Zhang(1993) 和 Shao(1993).

3.14.2 分离式变量选择

以整个协变量集 $\{x_1,\cdots,x_p\}$ 开始, 我们需要检查某一协变量 x_d 是否对响应变量 Y 有贡献. 为这个目的, 引入模型

$$Y = g(\beta^{\mathrm{T}}X_r, x_d) + e, \quad r\cup\{d\} = \{1,\cdots,p\}. \tag{3.14.5}$$

在模型 (3.1.1) 中, x_d 的贡献与其他协变量的贡献是通过线性组合混在一起的; 模型 (3.14.5) 中 x_d 的贡献是 "分离" 的, 且可以更精确地核定. 引入模型 (3.14.5) 的另一个理由是参数模型和非参数模型的交叉验证有不同的性质. 注意到在 (3.14.5) 中 Y 与 x_d 之间的相关性是 "非参数" 的. 因此, 正如 Cheng 和 Tong(1992) 与 Yao 和 Tong(1994) 已经证明的结论: leave-m-out 交叉验证可以告诉人们 x_d 是否对响应变量 Y 有显著贡献.

下面介绍 Kong 和 Xia(2007) 的方法. 对任何固定的 β, 定义 $g_{r,d}(x,v|\beta) = E(Y|\beta^{\mathrm{T}}X_r = \beta^{\mathrm{T}}x, x_d = v)$, 它的 leave-one-out 估计量 $\hat{g}_{r,d}^{\backslash i}(x,v|\beta)$ 为

$$\left\{\sum_{j\neq i} V_{j,r}(x,v|\beta)\Omega_{j,r}(x,v|\beta)\Omega_{j,r}^{\mathrm{T}}(x,v|\beta)\right\}^{-1}\sum_{j\neq i} V_{j,r}(x,v|\beta)\Omega_{j,r}^{\mathrm{T}}(x,v|\beta)Y_j, \tag{3.14.6}$$

其中 $\Omega_{j,r}(x,v|\beta) = (1, \beta^{\mathrm{T}}(X_{j,r}-x), x_{j,r}-v)^{\mathrm{T}}$, $V_{j,r}(x,v|\beta) = K_{h_1}(\beta^{\mathrm{T}}(X_{j,r}-x))H_{h_1}(x_{j,r}-v)$ 是二维的乘积核, h_1 为带宽; 如果 x_d 是连续的, $H = K$, 如果 x_d 是离散的, $H_{h_1}(v) = I(v=0)$.

为表达方便, 我们使用 $\hat{g}_{r_1,d}^{\backslash i}(X_i|\hat{\beta}_{r_1}^{\backslash i})$ 和 $\hat{g}_{r_1,d}^{\backslash i}(X_i|\hat{\beta}_{r_1}^{\backslash i})$ 分别表示 $\hat{g}_{r_1,d}^{\backslash i}(X_{i,r_1}, x_{i,d}|\hat{\beta}_{r_1}^{\backslash i})$ 和 $\hat{g}_{r_1,d}^{\backslash i}(X_{i,r_1}|\hat{\beta}_{r_1}^{\backslash i})$. 我们提出下列变量选择算法. 以一个满足 $r\supset r_0$ 的协变量集开始.

步骤 1. 在模型 $Y = g(\beta^{\mathrm{T}} X_r) + \varepsilon$ 下, 用所有数据点计算 β 的最小平均方差估计 $\hat{\beta}_r$. 用最小绝对值和在 r 中的相应指标 d 寻找 $\hat{\beta}_r$ 的表值. 令 $r_1 = r - \{d\}$.

步骤 2. 在模型 $Y = g(\beta^{\mathrm{T}} X_r) + \varepsilon$ 下, 基于 $\{(X_j, Y_j)\}_{j \neq i}$, 用 $\hat{\beta}_r^{\backslash i}$ 表示 β 的最小平均方差估计. 消除最小值并用 $\hat{\beta}_{r_1}^{\backslash i}$ 记剩余部分.

步骤 3. 用公式 (3.14.6) 计算 $\hat{g}_{r_1,d}^{\backslash i}(X_i|\hat{\beta}_{r_1}^{\backslash i})$, 用公式 (3.14.2) 计算 $\hat{g}_{r_1}^{\backslash i}(X_i|\hat{\beta}_{r_1}^{\backslash i})$, 其中 r 和 β 分别用 r_1 和 $\hat{\beta}_{r_1}^{\backslash i}$ 代替. 设

$$\mathrm{CV}_{r_1,d} = \frac{1}{n'} \sum_i{}' [Y_i - \hat{g}_{r_1,d}^{\backslash i}(X_i|\hat{\beta}_{r_1}^{\backslash i})]^2,$$

$$\mathrm{CV}_{r_1} = \frac{1}{n'} \sum_i{}' [Y_i - \hat{g}_{r_1}^{\backslash i}(X_i|\hat{\beta}_{r_1}^{\backslash i})]^2,$$

其中 $\sum_i{}'$ 表示对所有 $X_i \in A$ 的 i 求和. 这里 $A \subset \mathbf{R}^p$ 为紧凸集, 使得对所有接近于 β_0 的 β, $\beta^{\mathrm{T}} X$ 的密度函数在 $\{\beta^{\mathrm{T}} x : x \in A\}$ 上恒大于 0. 为方便起见, 假定: 如果 $1 \leqslant i \leqslant n'$, $X_i \in A$; 如果 $i > n'$, $X_i \bar{\in} A$.

步骤 4. 重复步骤 1~3 直到 $\mathrm{CV}_{r_1,d} < \mathrm{CV}_{r_1}$ 停止运算, 其运算用步骤 3 中的 r_1 代替步骤 1 中的 r. 最后选择 r.

称上面的程序为分离的交叉验证方法. 利用步骤 1 简化计算. 由于在单指标模型中 β_0 具有 $\sqrt{n}$ 相合估计, 所以若 $r \supset r_0$, 即 x_d 是多余的, 那么 $\hat{\beta}_d = O_P(n^{-1/2})$. 若 x_d 是必需的, 那么 $\hat{\beta}_d = \beta_{0,d} + O_P(n^{-1/2})$, 其以概率 1 有界地远离 0. 因此, 如果初始协变量集 r 包含多余变量, 那么对步骤 1 的排除仅以概率 1 考虑多余的变量. 用 $\hat{\beta}_r$ 和 $\hat{\beta}_{r_1}$ 分别代替 $\hat{\beta}_r^{\backslash i}$ 和 $\hat{\beta}_{r_1}^{\backslash i}$ 可以进一步简化步骤 2 和 3 的计算. 假定可以排除 x_d, 步骤 2 用来估计模型 (3.14.5) 中的参数. 步骤 3 计算和比较模型 (3.14.1) 和 (3.14.5) 的 leave-one-out 交叉验证值, 以便检查 x_d 的重要性; 见文献 (Cheng and Tong, 1992).

正如由 Härdle 等 (1993) 和 Xia(2006) 证明的结论, 非参数回归常用的带宽选择方法都可以用来估计联系函数和及指标系数. 当对 (3.14.6) 式计算时, 模型 (3.14.5) 的估计理应需要不同的带宽, 这依赖于 x_d 的类型: 若 x_d 连续 $h_1 \propto n^{-1/6}$, 若 x_d 离散 $h_1 = h \propto n^{-1/5}$, 其中 h 是计算 CV_{r_1} 所用的带宽. 许多带宽选择方法都可以用来选择这些带宽, 如交叉验证或广义交叉验证以及拇指法则等; 更详细的讨论可见文献 (Fan and Gijbels, 1996).

在一些正则条件下, Kong 和 Xia(2007) 证明了下列变量选择程序的相合性:

(i) 如果 $r \cup d = r_0$, 则 $\lim\limits_{n\to\infty} P\{\mathrm{CV}_{r,d} > \mathrm{CV}_r\} = 0$.

(ii) 如果 $r_0 \subseteq r$, $d \bar{\in} r_0$, 则 $\lim\limits_{n\to\infty} P\{\mathrm{CV}_{r,d} < \mathrm{CV}_r\} = 0$.

参考文献

Antoniadis A, Grégoire G, McKeague I W. 2004. Bayesian estimation in single-index models. Statistica Sinica, 14: 1147~1164

Arnold S F. 1981. The Theory of Linear Models and Multivariate Analysis. New York: John Wiley & Sons

Bai Y, Fung W K, Zhu Z Y. 2009. Penalized quadratic inference functions for singleindex models with longitudinal data. Journal of Multivariate Analysis, 100(1): 152~161

Berry S, Carroll R, Ruppert D. 2002. Bayesian smoothing and regression splines for measurement error problems. Journal of the American Statistical Association, 97: 160~169

Biller C. 2000. Adaptive Bayesian regression splines in semiparametric generalized linear models. Journal of Computational and Graphical Statistics, 9: 122~140

Brillinger D R. 1977. The identification of a particular nonlinear time series system. Biometrika, 64: 509~515

Brillinger D R. 1983. A generalized linear model with "gaussian" regressor variables. In A Festschrift for Erick L Lehmann Belmont CA: Wadsworth, 97~114

Carroll R J, Fan J, Gijbels I, et al.. 1997. Generalized partially linear singleindex models. Journal of the American Statistical Association, 92: 477~489

Casella G. 2001. Empirical Bayes Gibbs sampling. Biostatistics, 2: 485~500

Chang Z Q, Xue L G, Zhu L X. 2010. On an asymptotically more efficient estimation of the single-index Model. Journal of Multivariate Analysis, 101(8): 1898~1901

Chaudhuri P, Doksum K, Samarov A. 1997. On average derivative quantile regression. The Annals of Statistics, 25: 715~744

Cheng B, Tong H. 1992. On consistent nonparametric order determination and chaos (with discussion). Journal of the Royal Statistical Society, Series B, 54: 427~449

Chiou J M, Müller H G. 1998. Quasi-likelihood regression with unknown link and variance functions. Journal of the American Statistical Association, 93: 1376~1387

Choi T, Shi J Q, Wang B. 2011. A Gaussian process regression approach to a single-index model. Journal of Nonparametric Statistics, 23(1): 21~36

Cook R D. 1998. Regression Graphics: Ideas for Studying Regressions Through Graphics. New York: Wiley

Cook R D. 2007. Fisher lecture: dimension reduction in regression. Statistical Science, 22: 1~26

Cui X, Härdle W, Zhu L X. 2011. The EFM approach for single-index models. The Annals of Statistics, 39(3): 1658~1688

Delecroix M, Härdle W, Hristache M. 2003. Efficient estimation in conditional single-index regression. Journal of Multivariate Analysis, 862: 213~226

Delecroix M, Hristache M, Patilea V. 2006. On semiparametric M-estimation in single-index regression. Journal of Statistical Planning and Inference, 136(3): 730~769

Duan N, Li K C. 1991. Slicing regression: a link free regression method. The Annals of Statistics, 19: 505~530

Eilers P H C, Marx B D. 1996. Flexible smoothing with B-splines and penalties (with discussion). Statistical Science, 11: 89~121

Escanciano J C, Song K. 2010. Testing single-index restrictions with a focus on average derivatives. Journal of Econometrics, 156(2): 377~391

Fan J, Gijbels I. 1996. Local Polynomial Modeling and its Applications. London: Chapman & Hall

Gentle J E. 1998. Numerical Linear Algebra for Applications in Statistics. Berlin: Springer-Verlag

Green P. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika, 82: 711~732

Härdle W. 1990. Applied Nonparametric Regression. Cambridge: Cambridge University Press.

Härdle W, Hall P, Ichimura H. 1993. Optimal smoothing in single-index models. The Annals of Statistics, 21: 157~178

Härdle W, Mammen E. 1993. Testing parametric versus nonparametric regression. The Annals of Statistics, 21: 1926~1947

Härdle W, Spokoiny V, Sperlich S. 1997. Semiparametric single index versus fixed line function modelling. The Annals of Statistics, 25(1): 212~243

Härdle W, Stoker T M. 1989. Investigating smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association, 84: 986~995

Hädle W, Tsybakov A B. 1993. How sensitive are average derivative. Journal of Econometrics, 58: 31~48

Hoeffding W. 1963. Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association, 58: 13~30

Horowitz J L. 1998. Semiparametric Methods in Econometrics. London: Springer-Verlag

Horowitz J L, Härdle W. 1996. Direct semiparametric estimation of single-index models with discrete covariates. Journal of the American Statistical Association, 91: 1632~1640

Hristache M, Juditski A, Spokoiny V. 2001. Direct estimation of index coefficients in a single-index model. The Annals of Statistics, 29: 595~623

Hsing T, Carroll R J. 1992. An asymptotic theory for sliced inverse regression. The Annals of Statistics, 20: 1040~1061

Huang Z S, Zhang R Q. 2011. Efficient empirical-likelihood-based inferences for the single-index model. Journal of Multivariate Analysis, 102(5): 937~947

Huh J, Park B U. 2002. Likelihood-based local polynomial fitting for single-index models.

Journal of Multivariate Analysis, 80(2): 302～321
Ibragimov I, Khasminski R. 1987. Estimation if linear functionals in Gaussian noise. Theory of Probability and its Applications, 32: 30～39
Ichimura H. 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics, 58: 71～120
Jiang C R, Wang J L. 2011. Functional single index models for longitudinal data. The Annals of Statistics, 39(1): 362～388
Klein R, Shen C. 2010. Bias corrections in testing and estimating semiparametric single index models. Econometric Theory, 26(6): 1683～1718
Koenker R. 2005. Quantile Regression. New York: Cambridge University Press
Kohn R, Smith M, Chan D. 2001. Nonparametric regression using linear combinations of basis functions. Statistics and Computing, 11: 313～322
Kong E F, Xia Y C. 2007. Variable selection for the single-index model. Biometrika, 941: 217～229
Kong E F, Xia Y C. 2011. Quantile estimation of a general single-index model. Working paper, http://arxivorg/abs/08032474
Kulasekera K B, Lin W. 2010. Error variance estimation for the single-index model. Australian & new Zealand Journal of Statistics, 52(2): 201～219
Lee A J, Scott A J. 1986. Ultrasound in ante-natal diagnosis. The Fascination of Statistics, New York: Marcel Dekker, 277～293
Leitenstorfer F, Tutz G. 2011. Estimation of single-index models based on boosting techniques. Statistical Modelling, 11(3): 203～217
Li K C. 1989. Data visualization with SIR: a transformation based projection pursuit method. UCLA statistical series 24
Li K C. 1991. Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association, 86: 316～342
Lin W. 2011. On completely data-driven bandwidth selection for single-index models. Journal of Statistical Planning and Inference, 141(5): 1673～1683
Lin W, Kulasekera K B. 2007. Identitiability of single-index models and additive-index models. Biometrika, 94(2): 496～501
Lin W, Kulasekera K B. 2010. Testing the equality of linear single-index models. Journal of Multivariate Analysis, 101(5): 1156～1167
Lopez O. 2009. Single-index regression models with right-censored responses. Journal of Statistical Planning and Inference, 139(3): 1082～1097
Lu X W. 2010. Asymptotic distributions of two "synthetic data" estimators for censored single-index models. Journal of Multivariate Analysis, 101(4): 999～1015
Manski C F. 1988. Identification of binary response models. Journal of the American Statistical Association, 83: 729～738

McKeague I W, Tighiouart M. 2000. Bayesian estimators for conditional hazard functions. Biometrics, 56: 1007~1015

Naik P A, Tsai C L. 2000. Partial least squares estimator for single-index models. Journal of the Royal Statistical Society, Series B, 62: 763~771

Naik P A, Tsai C L. 2001. Single-index model selections. Biometrika, 88(3): 821~832

Naik P A, Tsai C L. 2004. Residual information criterion for single-index model selections. Journal of Nonparametric Statistics, 16(1-2): 187~195

Newey W K, Stoker T M. 1993. Efficiency of weighted average derivative estimators and index models. Econometrica, 61: 1199~1223

Ntzoufras I, Dellaportas P, Forster J J. 2003. Bayesian variable and link determination for generalised linear models. Journal of Statistical Planning and Inference, 111: 165~180

Ortega J M, Rheinboldt W C. 1973. Iterative Solution of Nonlinear Equations in Several Variables. New York: Academic Press.

Owen A B. 1988. Empirical likelihood ratio confidence intervals for a single function. Biometrika, 75: 237~249.

Owen A B. 1990. Empirical likelihood ratio confidence regions. The Annals of Statistics, 18: 90~120.

Peng H, Huang T. 2011. Penalized least squares for single index models. Journal of Statistical Planning and Inference, 141(4): 1362~1379

Petrov B B. 1991. Limit Theory of Sum of Independent Random Variable. New York: Springer-Verlag Inc.

Pollard D. 1984. Convergence of Stochastic Processes. New York: Springer-Verlag Inc.

Powell J L, Stock J H, Stoker T M. 1989. Semiparametric Estimation of Index Coefficients. Econometrica, 57: 1403~1430

Prendergast L A. 2005. Influence functions for sliced inverse regression. Scandinavian Journal of Statistics, 32(3): 385~404

Rao J N K, Scott A J. 1981. The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. Journal of the American Statistical Association, 76: 221~230

Samarov A. 1993. Exploring regression structure using nonparametric functional estimation. Journal of the American Statistical Association, 88: 836~847

Shao J. 1993. Linear model selection by cross-validation. Journal of the American Statistical Association, 88: 486~494

Smith M, Kohn R. 1996. Nonparametric regression using Bayesian variable selection. Journal of Econometrics, 75: 317~344

Stoker T M. 1986. Consistent Estimation of Scaled Coefficients. Econometrica, 546: 1461~1481

Stone M. 1974. Cross-validatory choice and assessment of statistical predictions (with

discussion). Journal of the Royal Statistical Society, Series B, 36: 111～147

Stute W, Zhu L X. 2005. Nonparametric checks for single-index models. The Annals of Statistics, 33(3): 1048～1083

Ulrich G. 1984. Computer generation of distributions on the m-sphere. Journal of the Royal Statistical Society, Series C, 33: 158～163

Wang H B. 2009. Bayesian estimation and variable selection for single index models. Computational Statistics and Data Analysis, 53: 2617～2627

Wang J L, Xue L G, Zhu L X, et al.. 2010a. Estimation for a partial-linear single-index model. The Annals of Statistics, 38(1): 246～274

Wang L, Yang L J. 2009. Spline estimation of single-index models. Statistica Sinica, 19(2): 765～783

Wang Y H, Shen J S, He S Y, et al.. 2010b. Estimation of single index model with missing response at random. Journal of Statistical Planning and Inference, 140(7): 1671～1690

Wang Y H, Li X Y, Wang Q H. 2011. Weighted estimation of single index models with right censored responses. Science China-Mathematics, 54(3): 479～514

Weisberg S, Welsh A H. 1994. Adapting for the missing link. The Annals of Statistics, 22(4): 1674～1700

Welsh A H. 1989. On M-processes and M-estimation. The Annals of Statistics, 17: 337～361. (Correction, 1990, 18: 1500)

Wu T Z, Yu K M, Yu Y. 2010. Single-index quantile regression. Journal of Multivariate Analysis, 101(7): 1607～1621

Xia Y C. 2006. Asymptotic distributions for two estimators of the single-index models. Econometric Theory, 22: 1112～1137

Xia Y C. 2007. A constructive approach to the estimation of dimension reduction directions. The Annals of Statistics, 35(6): 2654～2690

Xia Y C. 2008. A multiple-index model and dimension reduction. Journal of the American Statistical Association, 103: 1631～1640

Xia Y C. 2009. Model checking in regression via dimension reduction. Biometrika, 96(1): 133～148

Xia Y C, Li W K, Tong H, Zhang D X. 2004. A goodness-of-fit test for single-index models. Statistica Sinica, 14(1): 1～39

Xia Y C, Tong H, Li W K. 2002a. Single-index volatility models and estimation. Statistica Sinica, 12(3): 785～799

Xia Y C, Tong H, Li W K, Zhu L X. 2002b. An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society, Series B, 643: 363～410

Xue D, Xue L G, Cheng W H. 2011. Empirical likelihood for generalized linear models with missing responses. Journal of Statistical Planning and Inference, 141(6): 2007～2020

Xue L G. 2009a. Empirical likelihood confidence intervals for response mean with data missing at random. Scandinavian Journal of Statistics, 36(4): 671~685

Xue L G. 2009b. Empirical likelihood for linear models with missing responses. Journal of Multivariate Analysis, 100(7): 1353~1366

Xue L G. 2010. Empirical likelihood local polynomial regression analysis of clustered Data. Scandinavian Journal of Statistics, 37(4): 644~663

Xue L G, Xue D. 2011. Empirical likelihood for semiparametric regression model with missing response data. Journal of Multivariate Analysis, 102(4): 723~740

Xue L G, Zhu L X. 2006. Empirical likelihood for single-index models. Journal of Multivariate Analysis, 97: 1295~1312

Xue L G, Zhu L X. 2007a. Empirical likelihood for a varying coefficient model with longitudinal data. Journal of the American Statistical Association, 102(478): 642~654

Xue L G, Zhu L X. 2007b. Empirical likelihood semiparametric regression analysis for longitudinal data. Biometrika, 1~16

Yao Q, Tong H. 1994. On subset selection in nonparametric stochastic regression. Statist Sinica, 4: 51~70

Yin X R, Cook R D. 2005. Direction estimation in single-index regressions. Biometrika, 92: 371~384

Yu K, Lu Z. 2004. Local linear additive quantile regression. Scandinavian Journal of Statistics, 31: 333~346

Yu K, Jones M C. 1998. Local linear quantile regression. Journal of the American Statistical Association, 93: 228~237

Yu Y, Ruppert D. 2002. Penalized spline estimation for partially linear single-index models. Journal of the American Statistical Association, 97: 1042~1105

Yu Y, Ruppert D. 2004. Root-n consistency of penalized spline estimator for partially linear single-index models under general euclidean space. Statistica Sinica, 14(2): 449~455

Zhang P. 1993. Model selection via multifold cross validation. The Annals of Statistics, 21: 299~313

Zhang R Q, Huang Z S, Lv Y H. 2010. Statistical inference for the index parameter in single-index models. Journal of Multivariate Analysis, 101(4): 1026~1041

Zhou X H, Liang H. 2006. Semi-parametric single-index two-part regression models. Computational Statistics and Data Analysis, 50(5): 1378~1390

Zhu L P, Qian L Y, Lin J G. 2011. Variable selection in a class of single-index models. Annals of the institute of statistical mathematics, 63(6): 1277~1293

Zhu L P, Zhu L X. 2009. Nonconcave penalized inverse regression in single-index models with high dimensional predictors. Journal of Multivariate Analysis, 100(5): 862~875

Zhu L X, Fang K T. 1996. Asymptotics for the kernel estimates of sliced inverse regression.

The Annals of Statistics, 24: 1053~1067

Zhu L X, Ng K W. 1995. Asymptotics for sliced inverse regression. Statistica Sinica, 5: 727~736

Zhu L X, Xue L G. 2006. Empirical likelihood confidence regions in a partially linear single-index model. Journal of the Royal Statistical Society, Series B, 68: 549~570

第4章　部分线性单指标模型

4.1　引　　言

部分线性模型将传统的线性模型和非参数回归模型结合在一起，使得模型更具有灵活性，因而它倍受人们关注. 第 2 章已给出了详细的阐述. 然而，非参数分量常遭受维数灾祸且仅适应于低维协变量 X. 为弥补这一不足，人们假定受协变量 X 影响的降维模型可以作为一个单指标 $\beta_0^{\mathrm{T}}X$，并通过一个非参数函数 $g(\cdot)$ 使之与响应变量 Y 联系起来，这是切实可行的办法; 把它称为部分线性单指标模型. 假设 $\{(X_i,Z_i,Y_i),1\leqslant i\leqslant n\}$ 是来自 (X,Z,Y) 的 iid 样本，那么部分线性单指标模型具有形式

$$Y_i=\theta_0^{\mathrm{T}}Z_i+g(\beta_0^{\mathrm{T}}X_i)+\varepsilon_i,\quad i=1,\cdots,n, \tag{4.1.1}$$

其中 $g(\cdot)$ 为未知的联系函数，β_0 和 θ_0 分别为 p 维和 q 维未知参数，$(X_i,Z_i)\in\mathbf{R}^p\times\mathbf{R}^q$ 为响应变量 Y_i 的协变量，ε_i 为 iid 随机误差，且几乎处处 $E(\varepsilon_i|X_i,Z_i)=0$. 为了模型的可识别性，需要假定 $||\beta_0||=1$，并且 β_0 的第一个非 0 元素大于 0，这里 $||\cdot||$ 表示 Euclid 模.

模型 (4.1.1) 是相当一般的模型，它包含了前面提到的部分线性模型其 X 的维数是 1，以及单指标模型其缺少线性协变量 Z. 在第 3 章中，我们已详细阐述了单指标模型的估计方法. 现有的方法都依赖于预测 X_i 为连续变量的假定，而模型 (4.1.1) 允许有与响应变量发生线性联系的离散或其他连续变量而弥补了这一缺陷. 据我们所知，Carroll 等 (1997) 首次研究了模型 (4.1.1)，而他们实际上是考虑了一个广义的形式，称为广义部分线性单指标模型. 他们结合局部线性和最小二乘法构造了 (β_0,θ_0) 和 $g(\cdot)$ 的估计量，并得到了它们的渐近分布. 然而，正如 Yu 和 Ruppert(2002) 讲到的他们的方法在计算上不稳定. 此外，Carroll 等 (1997) 给出的定理也依赖于较强的假定: 他们的估计量已是 $\sqrt{n}$ 相合的. Yu 和 Ruppert(2002) 利用样条方法将原模型化成参数模型来缓解这两个麻烦. 他们用 P 样条估计联系函数，利用惩罚非线性最小二乘法来同时估计 P 样条单指标模型中的所有参数，该方法在计算上既快又稳定. 他们在假定紧的参数空间下证明了所有参数估计的 $\sqrt{n}$ 相合性和渐近正态性. 这些渐近结果说明了对参数的联合推断是可能的. Yu 和 Ruppert(2004) 在一般的参数空间上证明了同样的结果. Zou 等 (2009) 将 Cook 局部影响分析扩展到基于 P 样条的惩罚 Gauss 似然估计量. Leung 等 (2010) 使用光滑样条

方法研究了模型 (4.1.1). Xia 和 Härdle(2006) 基于局部线性光滑和 Härdle 等 (1993) 中最小二乘的改良形式提出了一个估计方法. Xia 等 (1999) 对模型 (4.1.1)$X = Z$ 的情形研究了有关参数的估计问题. Yin 和 Zhu(2007) 对广义部分线性单指标模型发展了一个估计方法, 所提出的方法可以推广到多指标模型来处理两个预测集.

Wang 等 (2010) 提出了两阶段估计方法来估计模型中的兴趣参数, 证明了参数分量的估计量的渐近正态性. 对指标系数, 一个估计方程得到了指标系数的一个更有效的估计, 因为它比其他现有估计量有更小的渐近方差. 联系函数的估计达到了最优的收敛速度. 此外, 他们也构造了误差方差的估计量并证明了它的渐近性质. 所得结果可以用来构造未知参数的置信域并可进行参数的大样本假设检验. Li 等 (2011) 纠正了 (Wang et al., 2010) 中引理 A.7 的错误. Liang 等 (2010) 构造了指标系数的半参数有效 profile 最小二乘估计量. 他们也利用 SCAD 方法同时选择变量和估计回归系数, 并且证明了提出的 SCAD 估计量是相合的且具有 oracle 性质, 同时也提出了一个调整参数的选择方法. 其他工作还有 (Zhao et al., 2011) 等.

Zhu 和 Xue(2006) 研究了部分线性单指标模型中参数的经验似然推断问题, 提出了 "纠偏" 的方法, 解决了参数经验似然比的构造问题. 不像现有的其他简单模型的经验似然方法, 如果没有纠偏, 经验似然比的 Wilks 定理不再成立, 见 3.10 节的讨论. 为了解决这个问题, Zhu 和 Xue(2006) 提出对经验似然比 "纠偏" 以达到标准 χ^2 分布. 纠偏的经验似然比有两个显著的特点: 当模型中有待估的无穷维讨厌函数时, 无需对联系函数欠光滑就可以确保兴趣参数估计的 $\sqrt{n}$ 相合性; 更为重要的是, 即便是有无穷维讨厌函数的嵌入估计量, 纠偏的经验似然仍是渐近标准 χ^2 的. 而在已有的文献中, 经验对数似然比的渐近分布是 χ^2 分布的加权和, 从而使统计推断变得复杂和麻烦. 此外, 使用指标向量的模为 1 这个信息提高了置信域的精度. 值得一提的是, 纠偏的思想可以应用到其他半参数回归模型中. Li 等 (2010) 在纵向数据下研究了部分线性单指标模型中参数的经验似然置信域的构造问题. Tian 等 (2010) 研究了纵向数据下部分线性单指标模型中参数估计的渐近性质.

Liang 和 Wang(2005) 研究了部分线性单指标 EV 模型. 他们构造了未知参数的两个估计量, 并证明了它们的渐近正态性. Chen 和 Cui(2009) 将经验似然方法应用到了部分线性单指标 EV 模型的研究, 证明了所构造的经验似然比的渐近 χ^2 性. Huang(2011a) 研究了部分线性单指标 EV 模型中参数分量的检验问题. 他构造了未知参数的广义似然比检验统计量并证明了相应的 Wilks 定理成立.

Lu 和 Cheng(2007) 研究了响应变量随机删失下部分线性单指标模型, 构造了模型中参数和非参数分量的估计量, 并证明了它们的渐近正态性. Lai 和 Wang(2011) 在缺失响应下研究了部分线性单指标模型的估计问题. 他们使用填补方法构造了模型中未知参数和联系函数的估计量, 并证明了这些估计量的渐近性质. 其他研究工作还包括 (He and Yi, 2011; Huang, 2011b; Liu et al., 2011) 等.

4.2 最小二乘估计和拟似然估计

在模型 (4.1.1) 下, 主要兴趣是估计 (β_0,θ_0) 和 $g(\cdot)$. 因为 $g(\cdot)$ 是非参数建模, 自然考虑局部线性光滑. 然而, 全局参数 β_0 和 θ_0 的有效估计需要使用所有数据, 因此它们的估计依赖于局部线性光滑. 对 u 的邻域内的点 v, 我们使用线性函数

$$g(v)\approx g(u)+g'(u)(v-u)\equiv a_0+a_1(v-u),$$

逼近 $g(\cdot)$, 其中 $a_0=g(u)$, $a_1=g'(u)$. 设 $K(\cdot)$ 为 $\mathbf{R}$ 上的核函数, $h=h_n$ 为趋于 0 的带宽序列. 记 $K_h(\cdot)=h^{-1}K(\cdot/h)$. 对固定的 β 和 θ, 极小化目标函数

$$L(a_0,a_1)=\sum_{i=1}^{n}[Y_i-\theta^{\mathrm{T}}Z_i-a_0-a_1(\beta^{\mathrm{T}}X_i-u)]^2K_h(\beta^{\mathrm{T}}X_i-u) \tag{4.2.1}$$

可得到它的极小值点 $\hat{a}_0$ 和 $\hat{a}_1$. 因此, $g(\cdot)$ 的局部线性估计为 $\hat{g}(u;\beta,\theta)=\hat{a}_0$. 简单计算可得

$$\hat{g}(u;\beta,\theta)=\sum_{i=1}^{n}W_{ni}(u,\beta)(Y_i-\theta^{\mathrm{T}}Z_i), \tag{4.2.2}$$

其中

$$W_{ni}(u;\beta)=\frac{n^{-1}K_h(\beta^{\mathrm{T}}X_i-u)[S_{n,2}(u;\beta,h)-(\beta^{\mathrm{T}}X_i-u)S_{n,1}(u;\beta,h)]}{S_{n,0}(u;\beta,h)S_{n,2}(u;\beta,h)-S_{n,1}^2(u;\beta,h)]}, \tag{4.2.3}$$

$$S_{n,l}(u;\beta,h)=\frac{1}{n}\sum_{i=1}^{n}(\beta^{\mathrm{T}}X_i-u)^lK_h(\beta^{\mathrm{T}}X_i-u),\quad l=0,1,2. \tag{4.2.4}$$

如果模型 (4.1.1) 是同方差的, 即 $\mathrm{var}(\varepsilon_i|X_i,Z_i)=\sigma^2$, 那么 β_0 和 θ_0 的估计为极小化问题

$$Q(\beta,\theta)=\frac{1}{n}\sum_{i=1}^{n}[Y_i-\theta^{\mathrm{T}}Z_i-\hat{g}(\beta^{\mathrm{T}}X_i;\beta,\theta)]^2 \tag{4.2.5}$$

的解. 称这个估计量为局部最小二乘估计. 如果模型(4.1.1)是异方差的, 即 $\mathrm{var}(\varepsilon_i|X_i,Z_i)=\sigma^2V(\theta_0^{\mathrm{T}}Z+g(\beta_0^{\mathrm{T}}X_i))$, 其中 $V(\cdot)$ 为某个已知函数, 那么 β_0 和 θ_0 的估计为极大化问题

$$R(\beta,\theta)=\frac{1}{n}\sum_{i=1}^{n}\int_{Y_i}^{\theta^{\mathrm{T}}Z_i+\hat{g}(\beta^{\mathrm{T}}X_i;\beta,\theta)}\frac{Y_i-u}{V(u)}\mathrm{d}u$$

的解. 称这个估计量为局部拟似然估计.

根据上述讨论, 可得到下列估计 β_0, θ_0 和 $g(\cdot)$ 的估计步骤.

步骤 0(初始化步). 拟合一个广义线性模型得到初始值 $\hat{\beta}$ 和 $\hat{\theta}$.

步骤 1. 极小化 $Q(\beta,\hat{\theta})$ 或极大化 $R(\beta,\hat{\theta})$ 得到 β_0 的模为 1 的估计 $\hat{\beta}$. 此时取 h 是一个估计的带宽, 使得 β_0 的估计达到最优.

步骤 2. 利用步骤 1 得到的估计 $\hat{\beta}$, 极小化 $Q(\hat{\beta},\theta)$ 或极大化 $R(\hat{\beta},\theta)$ 得到 θ_0 的估计 $\hat{\theta}$. 此时取 h 是一个估计的带宽, 使得 θ_0 的估计达到最优.

步骤 3. 继续步骤 1 和 2 直到收敛.

步骤 4. 利用 β_0 和 θ_0 的最终估计量 $\hat{\beta}$ 和 $\hat{\theta}$ 可得到 $g(u)$ 的估计量 $\hat{g}^*(u) = (u;\hat{\beta},\hat{\theta})$, 此时取 h 是一个估计的带宽, 使得当 β_0 和 θ_0 已知时 $g(\cdot)$ 的估计达到最优.

在步骤 1 中, 选择一个合理地带宽是很困难的. 一个相对合理地选择是

$$\hat{h}_{\text{opt}} \times n^{1/5} \times n^{-1/3} = \hat{h} \times n^{-2/15},$$

其中 $\hat{h}_{\text{opt}} \propto n^{-1/5}$ 是最优带宽. 这个带宽满足估计量 $\hat{\beta}$ 和 $\hat{\theta}$ 渐近正态性的条件. 在步骤 2 中, 可以选择最优带宽 $\hat{h}_{\text{opt}} \propto n^{-1/5}$, 这是部分线性模型所需要的带宽条件.

上述算法得到的估计量渐近等价于 Carroll 等 (1997) 给出的估计量. 这里我们对算法作了一点改进. Carroll 等 (1997) 证明了下列结果: 在一些正则条件下, 当带宽满足 $nh^2/\log n \to \infty$ 和 $nh^4 \to 0$ 时, 有

$$\sqrt{n}\begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{D} N(0, Q^-), \tag{4.2.6}$$

其中 Q^- 是 Q 的广义逆,

$$\begin{aligned} Q = E&\left[\begin{pmatrix} Xg'(U) \\ Z \end{pmatrix}\begin{pmatrix} Xg'(U) \\ Z \end{pmatrix}^{\mathrm{T}} \rho(\cdot)\right] \\ &- \begin{pmatrix} E[Xg'(U)\rho(\cdot)|U] \\ E[Z\rho(\cdot)|U] \end{pmatrix}\begin{pmatrix} E[Xg'(U)\rho(\cdot)|U] \\ E[Z\rho(\cdot)|U] \end{pmatrix}^{\mathrm{T}} \Big/ E[\rho(\cdot)|U], \end{aligned}$$

$U = \beta_0^{\mathrm{T}} X$, $\rho(\cdot) = [\sigma^2 V(\theta_0^{\mathrm{T}} Z + g(\beta_0^{\mathrm{T}} X))]^{-1}$.

特别地, 对同方差的情况, 即当 $\mathrm{var}(\varepsilon_i|X_i, Z_i) = \sigma^2$ 时, 有

$$Q = \sigma^{-2} E\left[\begin{pmatrix} Xg'(U) \\ Z \end{pmatrix}\begin{pmatrix} Xg'(U) \\ Z \end{pmatrix}^{\mathrm{T}}\right] - \begin{pmatrix} E[Xg'(U)|U] \\ E[Z|U] \end{pmatrix}\begin{pmatrix} E[Xg'(U)|U] \\ E[Z|U] \end{pmatrix}^{\mathrm{T}}.$$

更特别地, 如果没有 β_0 分量, 上述结果就是 Härdle 等 (1993) 在单指标模型下的相应结果.

4.3 惩罚样条估计

Yu 和 Ruppert(2002) 说明了惩罚样条估计方法提供了许多计算上的优势. 下面介绍这种方法在部分线性单指标模型中的应用. 假设

$$g(u) = \delta_0 + \delta_1 u + \cdots + \delta_d u_d + \sum_{i=1}^{k} \delta_{d+i}(u - s_i)_+^d, \tag{4.3.1}$$

其中 $\{s_i\}_{i=1}^k$ 是样条节点. 定义样条系数向量 $\delta = (\delta_0, \delta_1, \cdots, \delta_{d+k})^{\mathrm{T}}$ 和样条基

$$B(u) = (1, u, \cdots, u^d, (u - s_1)_+^d, \cdots, (u - s_k)_+^d)^{\mathrm{T}}.$$

那么 (4.3.1) 式可化为 $g(u) = \delta^{\mathrm{T}} B(u)$. 记 $\gamma = (\beta^{\mathrm{T}}, \theta^{\mathrm{T}}, \delta^{\mathrm{T}})^{\mathrm{T}}$. 极小化

$$Q_{n,\lambda}(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \theta^{\mathrm{T}} Z_i - \delta^{\mathrm{T}} B(\beta^{\mathrm{T}} X_i)\}^2 + \lambda \delta^{\mathrm{T}} D \delta$$

可得到 γ 的惩罚最小二乘估计, 记作 $\hat{\gamma}$. 其中 $\lambda \geqslant 0$ 为惩罚参数, D 为适当的半正定矩阵. 例如, D 可以时半正定矩阵使得

$$\delta^T D \delta = \int_{\min(\beta_0^{\mathrm{T}} X_i)}^{\max(\beta_0^{\mathrm{T}} X_i)} [g''(u)]^2 \mathrm{d}u,$$

其带来通常的二次积分惩罚. 如果 D 是对角矩阵, 其最后 k 个对角线上的元素为 1 其余全为 0. 它惩罚 $g(\cdot)$ 的第 d 阶跳跃的平方和. Yu 和 Ruppert(2002) 讨论了节点和 λ 的选择问题. 他们建议选 5~10 个节点就足够了, 且节点将由估计的指标值的等距量代替; 通过极小化 GCV 得分来选择惩罚参数 λ.

Yu 和 Ruppert(2002) 在参数空间是紧的场合和宽松的条件下证明了估计量 $\hat{\gamma}$ 的 $\sqrt{n}$ 相合性和渐近正态性. Yu 和 Ruppert(2004) 去掉了参数空间是紧的假定, 并在适当的条件下证明了惩罚最小二乘估计的存在性和 $\sqrt{n}$ 相合性.

4.4 最小平均方差估计

Xia 和 Härdle(2006) 考虑了部分线性单指标模型的高效、可构成和使用的估计量, 并将之应用到了时间序列之中. 他们提出的技术解决了来自 Carroll 等 (1997) 的两个问题: 一是不需要 $\sqrt{n}$ 相合的引导估计; 二是可由最优光滑速度选择光滑参数. 他们提出了模型中兴趣参数的最小平均方差估计量并获得了它们的渐近分布, 相应的算法容易被执行.

估计模型 (4.1.1) 中兴趣参数的基本算法是基于

$$(\beta_0, \theta_0) = \arg\min_{(\beta,\theta)} E\left[Y - \theta^{\mathrm{T}} Z - g(\beta^{\mathrm{T}} X)\right]^2, \quad \beta^{\mathrm{T}}\beta = 1. \tag{4.4.1}$$

在条件 $\xi = \beta^{\mathrm{T}} X$ 的作用下, 我们看到 (4.4.1) 等价于

$$(\beta_0, \theta_0) = \arg\min_{(\beta,\theta)} E\left[\sigma^2(\xi; \beta, \theta)\right], \quad \beta^{\mathrm{T}}\beta = 1, \tag{4.4.2}$$

其中

$$\sigma^2(\xi; \beta, \theta) = E\left\{[Y - \theta^{\mathrm{T}} Z - g(\beta^{\mathrm{T}} X)]^2 | \beta^{\mathrm{T}} X = \xi\right\}.$$

设 $\{(X_i, Z_i, Y_i), 1 \leqslant i \leqslant n\}$ 是来自 (X, Z, Y) 的样本. 可以用样本形式近似 (4.4.2) 式中的条件期望. 根据局部线性光滑的思想, 可以用下式估计 $\sigma^2(\beta^{\mathrm{T}} x; \beta, \theta)$:

$$\hat{\sigma}^2(\beta^{\mathrm{T}} x; \beta, \theta) = \min_{a,d} \sum_{i=1}^{n} \left[Y_i - \theta^{\mathrm{T}} Z_i - a - d\beta^{\mathrm{T}}(X_i - x)\right]^2 w_{i0}, \tag{4.4.3}$$

其中 $w_{i0} \geqslant 0 (1 \leqslant i \leqslant n)$ 是以 x 为中心的某个权, 且 $\sum_{i=1}^{n} w_{i0} = 1$. 记 $X_{ij} = X_i - X_j$. 提出的估计方法是针对 (a_j, d_j) 和 (β, θ) 极小化

$$\frac{1}{n} \sum_{j=1}^{n} G(\beta^{\mathrm{T}} X_j) I_n(X_j) \sum_{i=1}^{n} \left[Y_i - \theta^{\mathrm{T}} Z_i - a_j - d_j \beta^{\mathrm{T}} X_{ij}\right]^2 w_{ij} \tag{4.4.4}$$

其中 $G(\cdot)$ 是另一个权函数, 它是用来控制 (X_i, Z_i, Y_i) 对参数 β_0 和 θ_0 的贡献. 例如, 如果假定模型为异方差且 $\mathrm{var}(Y|X, Z) = V(\beta_0^{\mathrm{T}} X)$, 那么 $G(\cdot) = V(\cdot)$; 参见 (Carroll et al., 1997) 和 (Härdle et al., 1993). 利用 $I_n(x)$ 是出于技术的目的, 便于处理边界点. 也见 (Härdle et al., 1993). 为方便起见, 在实际中可以取 $I_n(\cdot) = 1$. 如果噪声项可忽略, 那么真参数 (β_0, θ_0) 是 (4.4.4) 式在 $w_{ij} \to 0$ 和 $X_i - X_j \to 0$ 下的极小化问题的解. 这个解释可以给我们一个直观的印象: 为什么极小化 (4.4.4) 式可以得到一个相合估计. 称这个估计程序为最小平均方差估计方法. 该方法与 Carroll 等 (1997) 和 Härdle 等 (1993) 的方法的主要差别是: 该方法是通过极小化 (4.4.4) 式的损失函数来估计所有参数和非参数函数, 而 Carroll 等 (1997) 和 Härdle 等 (1993) 给出的方法是通过极小化两个不同的损失函数来分别估计非参数函数和参数. 这个差别正是为什么该方法不需要欠光滑的原因. 类似 Härdle 等 (1993) 对单指标模型的估计思想, 其中他们对光滑参数 h 和指标系数 β_0 同时极小化一个常见的损失函数, 但对非参数函数是用另一个方法来估计.

通过固定 (a_j, d_j), $1 \leqslant j \leqslant n$ 和 (β, θ) 的其中任一个, 可以把极小化 (4.4.4) 式分成两类二次规划问题. 可以用一个解析表达式得到解决. 给定 (β, θ), 从极小化

(4.4.4) 式可得

$$\begin{pmatrix} a_j \\ d_j \end{pmatrix} = \left[\sum_{i=1}^{n} w_{ij} \begin{pmatrix} 1 \\ \beta^{\mathrm{T}} X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ \beta^{\mathrm{T}} X_{ij} \end{pmatrix}^{\mathrm{T}}\right]^{-1} \sum_{i=1}^{n} w_{ij} \begin{pmatrix} 1 \\ \beta^{\mathrm{T}} X_{ij} \end{pmatrix} (Y_i - \theta^{\mathrm{T}} Z_i) \tag{4.4.5}$$

给定 (a_j, d_j), 从极小化 (4.4.4) 式可得

$$\begin{aligned} \begin{pmatrix} \beta \\ \theta \end{pmatrix} = & \left[\sum_{j=1}^{n} G(\beta^{\mathrm{T}} X_j) I_n(X_j) \sum_{i=1}^{n} w_{ij} \begin{pmatrix} d_j X_{ij} \\ Z_i \end{pmatrix} \begin{pmatrix} d_j X_{ij} \\ Z_i \end{pmatrix}^{\mathrm{T}}\right]^{-1} \\ & \times \sum_{j=1}^{n} G(\beta^{\mathrm{T}} X_j) I_n(X_j) \sum_{i=1}^{n} w_{ij} \begin{pmatrix} d_j X_{ij} \\ Z_i \end{pmatrix} (Y_i - a_j) \end{aligned} \tag{4.4.6}$$

并标准化 $\beta := \mathrm{sgn}_1 \beta / \|\beta\|$, 其中 sgn_1 是 β 的第一个分量的符号. (4.4.4) 的极小化可以用 (4.4.5) 和 (4.4.6) 式之间的迭代来求解. 称上述方法为部分线性单指标算法.

权 w_{ij} 的选择在不同的估计方法中起着重要的作用. 见 (Hristache et al., 2001a, 2001b; Xia et al., 2002). 下面给出两种权. 设 $L(\cdot)$ 和 $K(\cdot)$ 分别是 p 个变量和单变量的对称密度函数. 第一种权是

$$w_{ij} = L_b(X_{ij}) \Big/ \sum_{k=1}^{n} L_b(X_{kj}),$$

其中 $L_b(\cdot) = b^{-1} L(\cdot/b)$, b 为带宽. 这是一个多维核权. 对这种权, 若 $n^{-1} \sum_{k=1}^{n} L_b(X_k - x) > c_0$, 那么置 $I_n(x) = 1$, 否则置 $I_n(x) = 0$, 其中 $c_0 > 0$ 为某个常数. 迭代 (4.4.5) 和 (4.4.6) 式直到收敛. 用 $\tilde{\beta}$ 和 $\tilde{\theta}$ 分别表示 β_0 和 θ_0 的估计量 (最终值). 因为 "维数灾祸", 所以基于这种权得到的估计量不是有效的. 然而, Xia 和 Härdle(2006) 证明了 $\tilde{\beta}$ 和 $\tilde{\theta}$ 分别是 β_0 和 θ_0 的相合估计. 因此, 多维核权可以帮助我们寻找 β_0 和 θ_0 的初始估计, 然后再使用单指标核权

$$w_{ij}^{\beta} = K_h(\beta^{\mathrm{T}} X_{ij}) \Big/ \sum_{k=1}^{n} K_h(\beta^{\mathrm{T}} X_{kj}),$$

其中 $K_h(\cdot) = h^{-1} K(\cdot/h)$, h 为带宽, β 取为 β_0 的估计 $\tilde{\beta}$. 这里若 $\hat{f}_{\beta}(\beta^{\mathrm{T}} x) > 0$, 取 $I_n(x) = 1$, 否则取 0, 其中

$$\hat{f}_{\beta}(u) = \frac{1}{n} \sum_{i=1}^{n} K_h(\beta^{\mathrm{T}} X_i - u). \tag{4.4.7}$$

迭代 (4.4.5) 和 (4.4.6) 式直到收敛. 用 $\hat{\beta}$ 和 $\hat{\theta}$ 分别表示 β_0 和 θ_0 的估计量 (最终值). 得到估计量 $\hat{\beta}$ 和 $\hat{\theta}$ 后, 再通过求解 (4.4.5) 就可以得到 $g(\cdot)$ 的估计. 用 $(\hat{a}_j, \hat{d}_j)$

表示 (4.4.5) 式的极小化问题的解, 那么 $\hat{g}(\beta^{\mathrm{T}}X_j;\beta,\theta)=\hat{a}_j$, $\hat{g}'(\beta^{\mathrm{T}}X_j;\beta,\theta)=\hat{d}_j$. 因此, $g(u)$ 的估计量为 $\hat{g}^*(u)=\hat{g}(u;\hat{\beta},\hat{\theta})$.

记 $\mu_\beta(u)=E(X|\beta^{\mathrm{T}}X=u)$, $\nu_\beta(u)=E(Z|\beta^{\mathrm{T}}X=u)$, $k=0,2$,

$$W_k=E\left\{G(\beta_0^{\mathrm{T}}X)I\{f_{\beta_0}(\beta_0^{\mathrm{T}}X)>c_0\}\begin{pmatrix}g'(\beta_0^{\mathrm{T}}X)[X-\mu_{\beta_0}(\beta_0^{\mathrm{T}}X)]\\Z-\nu_{\beta_0}(\beta_0^{\mathrm{T}}X)\end{pmatrix}\right.$$
$$\left.\times\begin{pmatrix}g'(\beta_0^{\mathrm{T}}X)[X-\mu_{\beta_0}(\beta_0^{\mathrm{T}}X)]\\Z-\nu_{\beta_0}(\beta_0^{\mathrm{T}}X)\end{pmatrix}^{\mathrm{T}}\varepsilon^k\right\}.$$

Xia 和 Härdle(2006) 在样本为强混合平稳序列下证明了 $\hat{\beta}$ 和 $\hat{\theta}$ 的渐近正态性. 他们给出了下列结果: 在适当的正则条件下, 如果 $h\sim n^{-\delta}$, $1/6<\delta<1/4$, 且几乎处处有 $E[\varepsilon_i|X_j,Z_j,Y_j,J<i]=0$, 则

$$\sqrt{n}\begin{pmatrix}\hat{\beta}-\beta_0\\\hat{\theta}-\theta_0\end{pmatrix}\xrightarrow{D}N(0,W_0^-W_2W_0^-),\tag{4.4.8}$$

其中 W_0^- 是 W_0 的 Moore-Penrose 逆. 如果进一步假定 $\beta_0^{\mathrm{T}}X$ 的密度 $f_{\beta_0}(u)$ 是正的, 且 $\gamma^2(u)=E(\varepsilon^2|\beta_0^{\mathrm{T}}X=u)$ 的导数存在, 那么

$$\sqrt{nh}[\hat{g}^*(u)-g(u)-\kappa_2g''(u)h^2/2]\xrightarrow{D}N(0,\nu_0f_{\beta_0}^{-1}(u)\gamma^2(u)),\tag{4.4.9}$$

其中 $\nu_0=\displaystyle\int K^2(u)\mathrm{d}u$, $\kappa_2=\displaystyle\int u^2K(u)\mathrm{d}u$.

如果 $E[\varepsilon_i|X_j,Z_j,Y_j,J<i]\neq0$, 那么 $\{\hat{\beta},\hat{\theta}\}$ 的渐近正态性仍成立, 但分布中的协方差阵依赖于观测的随机过程的分布. 如果 $E(\varepsilon^2|X,Z)=\sigma^2$ 是常数, 那么 $\{\hat{\beta},\hat{\theta}\}$ 的渐近分布与 Carroll 等 (1997) 得到的渐近分布相同. 因此, 在同样的条件及半参数意义下, 这里的估计同样有效. 如果要使用 (4.4.8) 式对参数作统计推断, 需要构造渐近方差的相合估计. 这里给出 W_k 的相合估计:

$$\hat{W}_k=\frac{1}{n}\sum_{i=1}^{n}G(\hat{\beta}^{\mathrm{T}}X_i)I\{\hat{f}_{\hat{\beta}}(\hat{\beta}^{\mathrm{T}}X_i)>c_0\}\begin{pmatrix}\hat{g}'(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta})[X_i-\hat{\mu}_{\hat{\beta}}(\hat{\beta}^{\mathrm{T}}X_i)]\\Z_i-\hat{\nu}_{\hat{\beta}}(\hat{\beta}^{\mathrm{T}}X_i)\end{pmatrix}$$
$$\times\begin{pmatrix}\hat{g}'(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta})[X_i-\hat{\mu}_{\hat{\beta}}(\hat{\beta}^{\mathrm{T}}X_i)]\\Z_i-\hat{\nu}_{\hat{\beta}}(\hat{\beta}^{\mathrm{T}}X_i)\end{pmatrix}^{\mathrm{T}}[Y_i-\hat{\beta}^{\mathrm{T}}Z_i-\hat{g}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta})]^k,$$

其中 $k=0,2$, $\hat{f}_{\hat{\beta}}(u)$ 在 (4.4.7) 式中定义,

$$\hat{\mu}_{\hat{\beta}}(u)=\frac{1}{n}\sum_{i=1}^{n}K_h(\hat{\beta}^{\mathrm{T}}X_i-u)X_i/\hat{f}_{\hat{\beta}}(u),$$

$$\hat{\nu}_{\hat{\beta}}(u)=\frac{1}{n}\sum_{i=1}^{n}K_h(\hat{\beta}^{\mathrm{T}}X_i-u)Z_i/\hat{f}_{\hat{\beta}}(u).$$

带宽选择总是非参数光滑的重要问题. 本节方法的优点之一是不需要欠光滑联系函数. 因此, 这里可以使用大多数通常使用的带宽选择方法. 考虑最后迭代步得到的估计量 $\hat{g}^*(u)$. 对给定的具有紧支撑的函数 $w(\cdot)$, 极小化具有权 $f_{\beta_0}(\cdot)w(\cdot)$ 的渐近加权均方误差可得到最优整体带宽

$$h_0 = \left[\frac{\nu_0\sigma^2 \int w(u)\mathrm{d}u}{\kappa_2^2 \int g''(u)f_{\theta_0}(u)w(u)\mathrm{d}u}\right]^{1/5} n^{-1/5}.$$

见 Xia 和 Härdle(2006) 与 Carroll 等 (1997) 的讨论. 因此, 可以利用交叉核实法选择最优带宽. 对任何 β 和 θ, 令

$$\mathrm{CV}_0(b) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \theta^{\mathrm{T}}Z_i - \tilde{a}_i^{\setminus i})^2,$$

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \theta^{\mathrm{T}}Z_i - \hat{a}_i^{\setminus i})^2,$$

其中 $\tilde{a}_i^{\setminus i}$ 和 $\hat{a}_i^{\setminus i}$ 是删除第 i 个观测 (X_i, Z_i, Y_i) 后利用公式 (4.4.5) 得到估计量, 其中 $\tilde{a}_i^{\setminus i}$ 中使用权函数 w_{ij}, 而 $\hat{a}_i^{\setminus i}$ 中使用权函数 w_{ij}^{β}. 因此, 可得到每一步使用的带宽: $\hat{b} = \arg\min_b \mathrm{CV}_0(b)$, $\hat{h} = \arg\min_h \mathrm{CV}_0(h)$.

4.5 估计方程估计

本节将局部线性光滑和最小二乘法相结合构造了模型 (4.1.1) 中兴趣参数的估计方程估计, 然后给出估计量的渐近性质. 本节的主要内容来自 (Wang et al., 2010; Li et al., 2011).

4.5.1 估计方法

本节假定模型 (4.1.1) 为同方差的, 即 $\mathrm{var}(\varepsilon|X,Z) = \sigma^2 > 0$. 下面提出了一个新的估计方法. 这个方法不需要迭代且在宽松的条件下具有优良效果, 其条件是基于 X_i 的几个指标解释 Z_i, 即

$$Z_i = \phi(\beta_Z^{\mathrm{T}}X_i) + \eta_i, \quad i = 1, \cdots, n, \tag{4.5.1}$$

其中 $\phi(\cdot)$ 是 $\mathbf{R}^d$ 到 $\mathbf{R}^q$ 上的未知函数, β_Z 为具有正交列的 $p \times d$ 矩阵, η_i 具有均值 0 且独立于 X_i. 维数 d 通常小于 X_i 的维数 p. 这样一个假定不是很强且常常用在降维方法的文献中. Li 等 (2008) 提供了它的理论根据. 模型 (4.5.1) 蕴涵 X_i 的几

个指标可足以包含 X_i 输入给预测 Z_i 的所有信息, 这种情况在实际中经常发生, 见 (Wang et al., 2010).

4.2 节中提供的方法是同时估计 β_0 和 θ_0, 其算法带来计算上的复杂度. 为避免这个不足, 我们建议分别估计 β_0 和 θ_0. 这个思想很简单：一旦有 β_0 的一个 $\sqrt{n}$ 相合估计, 就可以嵌入到模型 (4.1.1) 中, 然后利用部分线性模型的估计方法来估计 θ_0. 然而, β_0 和 θ_0 可能是相关的, 这就带来识别 β_0 的困难. 此时 (4.5.1) 式将为统计推断提供帮助, 因为它可以去掉 Z_i 相关于 X_i 的部分, 使得 (4.5.1) 式中的残差 η_i 独立于 X_i. 此外, 我们需要强加可识别性条件: β_Z 具有模 1 且第一个分量是正的. 本节的估计思想如下: 首先利用任何一个降维方法估计 β_Z. 然后利用 d 维光滑方法估计 $\phi(\cdot)$, 从而得到残差 η_i. 因为 $\eta_i = Z_i - \phi(\beta_Z^{\mathrm{T}} X_i)$, 将它嵌入 (4.1.1) 式中, 可得

$$Y_i = \theta_0^{\mathrm{T}} \eta_i + \psi(\beta_0^{\mathrm{T}} X_i, \beta_Z^{\mathrm{T}} X_i) + \epsilon_i, \quad i = 1, \cdots, n,$$

其中 $\psi(\cdot)$ 是一个未知函数, 但现在 η_i 和 X_i 是相互独立的. 因此可以利用 Y_i 关于 η_i 的线性模型的最小二乘方法估计 θ_0, 且所得估计 $\hat{\theta}_0$ 是 $\sqrt{n}$ 相合的. 然后对 $Y_i - \hat{\theta}_0^{\mathrm{T}} Z_i$ 和 X_i 利用降维方法得到 β_0 和 $g(\cdot)$ 的估计. 这就结束了第一阶段, 其中所得到的 θ_0 和 β_0 的估计已是 $\sqrt{n}$ 相合的, 但是它们将扮演为下一步估计提供初始值的角色. 第二步使用更优良的方法更新所有估计. 具体地, 利用偏回归方法估计 θ_0, 利用估计方程方法估计 β_0, 从而得到比现有估计量更有效的估计. 下一节的理论结果说明两阶段估计步骤完全有效, 当然也不需要迭代.

为表达方便, 我们最初假定 Z_i 与 X_i 的单指标相关, 即在 (4.5.1) 式中 $d = 1$. $d > 1$ 的情况在注 4.5.1 中给出了说明. 下面首先概述估计步骤, 然后再对每一步展开阐述.

第一阶段的算法:

(1) 应用 Z_i 对 X_i 回归的降维方法找 β_Z 的估计 $\hat{\beta}_Z$;

(2) 利用数据 $\{(\hat{\beta}_Z^{\mathrm{T}} X_i, Z_i), 1 \leqslant i \leqslant n\}$ 和光滑方法得到 $\phi(\cdot)$ 的估计 $\hat{\phi}(\cdot)$, 然后计算残差 $\hat{\eta}_i = Z_i - \hat{\phi}(\hat{\beta}_Z^{\mathrm{T}} X_i; \hat{\beta}_Z)$;

(3) 运用 Y_i 对 $\hat{\eta}_i$ 的线性回归找 θ_0 的初始估计 $\hat{\theta}_0$;

(4) 将降维方法应用到 $Y_i - \hat{\theta}_0^T Z_i$ 对 X_i 的单指标回归来找 β_0 的一个初始估计 $\hat{\beta}_0$;

(5) 用光滑方法和数据 $\{(\hat{\beta}_0^T X_i, Y_i - \hat{\theta}_0^T Z_i), 1 \leqslant i \leqslant n\}$ 得到 $g(\cdot)$ 及其导数 $g'(\cdot)$ 的估计;

第二阶段的算法:

(6) 使用步骤 (4) 的初始估计 $\hat{\beta}_0$ 和部分线性模型的 profile 方法来更新 θ_0 的估计, 即极小化 (4.5.3) 式得到 θ_0 的估计 $\hat{\theta}$;

(7) 使用步骤 (6) 的更新估计 $\hat{\theta}$ 形成新的残差 $Y_i-\hat{\theta}^{\mathrm{T}}Z_i$, 然后解估计方程 (4.5.6) 来更新 β_0 的估计;

(8) 使用步骤 (6) 和 (7) 中 θ_0 和 β_0 的更新估计并按照步骤 (5) 的方法来更新 $g(\cdot)$ 的估计.

这就完成了算法的描述. 虽然下一节证明了所得到得估计量在理论上是有效的, 但可通过重复步骤 (5) 和 (6) 一次或更多次来改进实际计算的精度. 模拟计算表明: 当重复次数大于 1 时其计算精度提高甚微.

下面我们就单指标的简单情况 ($d=1$) 详细说明上面算法的每一步. 对步骤 (4) 的降维方法, 人们可以使用现有方法的任何一种来得到初始估计 $\hat{\beta}_0$, 如, SIR 和它的变形, Xia 等 (2002) 的最小平均导数方法等. 这些方法都是对单变量响应, 因此当 $q=1$ 时可以在步骤 1 中应用. 然而, 当 $q>1$ 时, 对多变量响应步骤 (1) 需要用不同的方法, 我们推荐 Li 等 (2008) 的降维方法. 文献中的相关结果已经证明这些降维方法所得到的估计量是 $\sqrt{n}$ 相合的.

对参与在步骤 (5) 中的光滑, 人们可以选择任何一维光滑. 我们利用局部线性光滑 (Fan and Gijbels, 1996) 得到联系函数 $g(u)$ 及其导数 $g'(u)$ 的估计量. 具体地, 极小化目标函数 (4.2.1) 可得到 $g(u)$ 和 $g'(u)$ 的估计量, 并记为 $\hat{g}(u;\beta,\theta)$ 和 $\hat{g}'(u;\beta,\theta)$, 其中 $\hat{g}(u;\beta,\theta)$ 已在 (4.2.2) 式中定义, 而

$$\hat{g}'(u;\beta,\theta)=\sum_{i=1}^{n}\widetilde{W}_{ni}(u,\beta)(Y_i-Z_i^{\mathrm{T}}\theta). \tag{4.5.2}$$

这里

$$\widetilde{W}_{ni}(u;\beta)=\frac{n^{-1}K_{h_1}(\beta^{\mathrm{T}}X_i-t)[(\beta^{\mathrm{T}}X_i-t)S_{n,0}(u;\beta,h_1)-S_{n,1}(u;\beta,h_1)]}{S_{n,0}(u;\beta,h_1)S_{n,2}(u;\beta,h_1)-S_{n,1}^2(u;\beta,h_1)]},$$

$S_{n,l}(u;\beta,h)$ 在 (4.2.4) 中定义, $h_1=h_{1n}$ 为趋于 0 的带宽序列, $K_{h_1}(\cdot)=h_1^{-1}K(\cdot/h_1)$, $K(\cdot)$ 为 $\boldsymbol{R}$ 上的核函数.

上述估计量是对 β 和 θ 的一个固定值. 分别用步骤 (1) 和 (3) 得到的初始值 $\hat{\beta}_0$ 和 $\hat{\theta}_0$ 代替它们, 就得到步骤 (5) 所需要的估计量. 下面的定理 4.5.3 证明了它导致 $g(u)$ 的估计量的标准收敛速度.

类似地, 为估计模型 (4.5.1) 中的 $\phi(\cdot)$, 也可以在步骤 (2) 中利用局部线性光滑. 所得到的 $\phi(\cdot)$ 的估计量定义为

$$\hat{\phi}(u;\hat{\beta}_Z)=\sum_{i=1}^{n}W_{ni}(u;\hat{\beta}_Z)Z_i,$$

其中 $W_{ni}(u;\hat{\beta}_Z)$ 在 (4.2.3) 式中定义.

可以利用多种方法得到步骤 6 中 θ_0 的估计量, 比如, profile 方法 (也称偏回归, Speckman, 1988) 或部分样条 (Heckman, 1986). 这里部分样条方法不适合相关的 X 和 Z, 因此采取 profile 方法和局部线性光滑. 简言之, 这相当于极小化 θ 的目标函数

$$\sum_{i=1}^{n}[Y_i - \theta^{\mathrm{T}} Z_i - \hat{g}(\hat{\beta}_0^{\mathrm{T}} X_i; \hat{\beta}_0, \theta)]^2, \tag{4.5.3}$$

其中 $\hat{g}(\cdot; \hat{\beta}_0, \theta)$ 是由 (4.2.2) 式定义的 $g(\cdot)$ 的估计量, $\hat{\beta}_0$ 是 β_0 的估计量, 它可以是步骤 (4) 中的初始估计 $\hat{\beta}_0$, 也可以是来自步骤 (7) 的估计 (此时得到 θ_0 的迭代估计). 因为光滑器是 θ 的函数, 所以此估计由 (4.2.3) 式获得. 关于 profile 方法的来历和优点可参见 (Speckman, 1988). 具体地, 记 $\hat{\beta}_0$ 是现行的估计量, $\widetilde{\boldsymbol{Y}} = (\widetilde{Y}_1, \cdots, \widetilde{Y}_n)^{\mathrm{T}}$, $\widetilde{\boldsymbol{Z}} = (\widetilde{Z}_1, \cdots, \widetilde{Z}_n)^{\mathrm{T}}$, 其中

$$\widetilde{Y}_i = Y_i - \hat{g}_1(\hat{\beta}_0^{\mathrm{T}} X_i; \hat{\beta}_0), \qquad \widetilde{Z}_i = Z_i - \hat{\nu}(\hat{\beta}_0^{\mathrm{T}} X_i; \hat{\beta}_0),$$

$$\hat{g}_1(u; \hat{\beta}_0) = \sum_{i=1}^{n} W_{ni}(u; \hat{\beta}_0) Y_i, \quad \hat{\nu}(u; \hat{\beta}_0) = \sum_{i=1}^{n} W_{ni}(u; \hat{\beta}_0) Z_i,$$

$\hat{g}_1(u; \hat{\beta}_0)$ 和 $\hat{\nu}(u; \hat{\beta}_0)$ 分别是 $g_1(u) = E(Y|X^{\mathrm{T}}\beta_0 = u)$ 和 $\nu(u) = E(Z|X^{\mathrm{T}}\beta_0 = u)$ 的估计量, 其中 $W_{ni}(u; \hat{\beta}_0)$ 在 (4.2.3) 式中定义. 因此, 偏回归估计为

$$\hat{\theta} = (\widetilde{\boldsymbol{Z}}^{\mathrm{T}} \widetilde{\boldsymbol{Z}})^{-1} \widetilde{\boldsymbol{Z}}^{\mathrm{T}} \widetilde{\boldsymbol{Y}}. \tag{4.5.4}$$

为在步骤 (7) 中估计 β_0, 我们利用约束 $||\beta_0|| = 1$ 提出了一个新的估计方法; 此方法所得到的估计比现有方法得到的估计更有效, 包括 Härdle 等 (1993) 与 Carroll 等 (1997) 的最小二乘法以及 Xia 等 (2002) 与 Xia 和 Härdle(2006) 的最小平均方差估计法. 值得一提的是, Xia 和 Härdle(2006) 允许模型 (4.1.1) 为异方差结构, 且得到 β_0 的估计量的同样的渐近方差. 例如, 在同方差的情况下, Xia 和 Härdle(2006) 得到的估计量的渐近方差与 Härdle 等 (1993) 得到的渐近方差相同. 本节提出的估计方程方法计算上稳定且更有效, 即它的渐近方差更小. 有效性的获得归功于使用约束 $||\beta|| = 1$ 对 β 进行再参数化, 把一个约束最小二乘问题转化为无约束最小二乘问题, 使得在 Euclid 空间 $\mathbf{R}^{p-1}$ 的一个约束区域内尽可能搜寻估计方程的所有解.

为导出估计方程, 我们利用最小二乘准则:

$$D(\beta) = \frac{1}{n} \sum_{i=1}^{n} [Y_i - \hat{\theta}^{\mathrm{T}} Z_i - \hat{g}(\beta^{\mathrm{T}} X_i; \beta, \hat{\theta})]^2 w(\beta^{\mathrm{T}} X_i), \tag{4.5.5}$$

其中 $w(\cdot)$ 是一个具有有界支撑 $\mathcal{U}_w$ 的有界权函数, 它是用来控制估计量 $\hat{g}(\cdot)$ 的边界效应, 这里取 $w(\cdot)$ 为 $\mathcal{U}_w$ 上的示性函数. 极小化 $D(\beta)$ 可得到 β_0 的估计量 $\hat{\beta}$. 根

据 3.10 节的讨论可得到估计方程组

$$\begin{cases} \sum_{i=1}^{n}[Y_i - \hat{\theta}^{\mathrm{T}} Z_i - \hat{g}(\beta^{\mathrm{T}} X_i; \beta, \hat{\theta})]\hat{g}'(\beta^{\mathrm{T}} X_i; \beta, \hat{\theta}) J_{\beta^{(r)}}^{\mathrm{T}} X_i w(\beta^{\mathrm{T}} X_i) = 0, \\ ||\beta|| - 1 = 0, \end{cases} \tag{4.5.6}$$

其中 $J_{\beta^{(r)}}$ 在 (3.10.3) 式中定义. 解方程组 (4.5.6) 即可得到 β_0 的估计量 $\hat{\beta}$.

利用估计量 $\hat{\theta}$ 和 $\hat{\beta}$, 可定义步骤 (8) 中 $g(u)$ 的估计量

$$\hat{g}^*(u) \equiv \hat{g}(u; \hat{\beta}, \hat{\theta}) = \sum_{i=1}^{n} W_{ni}(u; \hat{\beta})(Y_i - \hat{\theta}^{\mathrm{T}} Z_i), \tag{4.5.7}$$

σ^2 的估计量定义为

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}[Y_i - \hat{\theta}^{\mathrm{T}} Z_i - \hat{g}^*(\hat{\beta}^{\mathrm{T}} X_i)]^2. \tag{4.5.8}$$

下面的定理 4.5.1 和定理 4.5.2 分别给出了估计量 $\hat{\beta}$ 和 $\hat{\theta}$ 的渐近正态性; 定理 4.5.3 给出了估计量 $\hat{g}^*(u)$ 的收敛速度; 定理 4.5.4 给出了估计量 $\hat{\sigma}^2$ 的渐近正态性.

注 4.5.1 到目前为止, 我们假定 $d = 1$. 这个假定可以容易地推广到 $d > 1$. 在这种情况下, 为估计 $\phi(\cdot)$ 需要利用多维光滑. 此时, 除了 $\phi(\cdot)$ 的估计量的收敛速度随着维数 d 的改变而改变外, β_0 和 θ_0 的估计量的渐近结果仍不变.

注 4.5.2 当 Z 的维数 q 大于 1 时, 算法的步骤 (1) 原则上可以利用 SIR(Li et al., 2003) 的多维扩展. 然而, 每片观察的数目可能变得稀少, 因此我们推荐诸如 Li 等 (2008) 的多维方法.

注 4.5.3 本节对模型 (4.1.1) 的单指标假定可以通过 SIR 及其变形容易推广到多指标情形, 但多维联系函数的估计将会遭到维数灾祸. 因为在许多实际问题中不需要大于 3 个指标, 因此本节的方法的确可以在实际中推广到多维情形.

4.5.2 主要结果

在本节中, 第一阶段 β_0 和 θ_0 的初始估计的 $\sqrt{n}$ 相合性已认为来自现有结果, 因此我们没有列出它们所需要的条件, 但在下面定理 4.5.1 之后提供了这些结果成立的可能性. 为研究最终估计量的渐近性质, 我们列出下列条件.

(C4.5.1) (i) $\beta_0^{\mathrm{T}} X$ 的密度函数 $f(u)$ 在 $\mathcal{U}_w$ 上一致大于 0, 且满足 1 阶 Lipschitz 条件, 其中 $\mathcal{U}_w$ 是 $w(u)$ 的支撑集.

(ii) 对 β_0 邻域内的点 β, $\beta^{\mathrm{T}} X$ 的密度函数是正的, 且满足 1 阶 Lipschitz 条件.

(C4.5.2) (i) $g(u)$ 和 $\nu(u)$ 的每一个分量具有连续有界的二阶导数.

(ii) $\mu(u) = E(X|\beta_0^{\mathrm{T}}X = u)$ 的每一个分量满足 1 阶 Lipschitz 条件.

(C4.5.3) (i) 核 $K(u)$ 是有界、连续且对称的概率密度函数, 且满足 $0 < \int_{-\infty}^{\infty} u^2K(u)\mathrm{d}u < \infty$.

(ii) $K(u)$ 在 $\boldsymbol{R}$ 上满足 Lipschitz 条件.

(C4.5.4) (i) $\sup_u E(||X||^4|\beta_0^{\mathrm{T}}X = u) < \infty$, $\sup_u E(||Z||^2|\beta_0^{\mathrm{T}}X = u) < \infty$.

(ii) $0 < \mathrm{var}(\varepsilon|X) = \sigma^2 < \infty$, $\sup_{x,z} E(\varepsilon^4|X = x, Z = z) < \infty$.

(C4.5.5) (i) $nh^2/\log^2 n \to \infty$, $\limsup_{n\to\infty} nh^5 < \infty$.

(ii) $nhh_1^3/\log^2 n \to \infty$, $nh^4 \to 0$, $\limsup_{n\to\infty} nh_1^5 < \infty$.

(C4.5.6) $\Sigma = \mathrm{cov}(Z - E(Z|\beta_0^{\mathrm{T}}X))$ 是正定矩阵.

关于对条件的注释可参见 (3.10.1). 下列定理陈述了上节提出的估计量的渐近性质. 首先建立 $\hat{\theta}$ 的渐近性质.

定理 4.5.1 假设条件 (C4.5.1)~(C4.5.6) 成立. 如果 $||\hat{\beta}_Z - \beta_Z|| = O_P(n^{-1/2})$ 和 $||\hat{\beta}_0 - \beta_0|| = O_P(n^{-1/2})$, 那么

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, \sigma^2\Sigma^{-1}),$$

其中 Σ 在条件 (C4.5.6) 中定义.

在定理 4.5.1 中, 当开始于 β_Z 和 β_0 的 $\sqrt{n}$ 相合估计时, $\hat{\theta}$ 是 θ_0 的渐近有效估计. 在文献中已经存在 $\sqrt{n}$ 相合估计的许多例子. 例如, 在比 X 的椭球对称稍弱的线性条件下, Li(1991), Hsing 和 Carroll(1992) 与 Zhu 和 Ng(1995) 证明了 SIR 导致 β_Z 和 β_0 的 $\sqrt{n}$ 相合估计量. Xia 等 (2002) 提出的最小平均方差估计方法也能够提供 β_0 的 $\sqrt{n}$ 相合估计. 然而, 当 Z 的维数大于 1 时还没有最小平均方差估计的类似结果, 且当 $d \geqslant 3$ 时需要进一步研究相应估计量的 $\sqrt{n}$ 相合性, 即使对单变量 Z 也是如此. 因此, 对一般定理, 使用 SIR 得到 β_Z 和 β_0 的初始估计或许是一种好的选择.

定理 4.5.2 假设条件 (C4.5.1)~(C4.5.6) 成立. 如果 β_0 的第 r 个分量是正的, 那么

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, \sigma^2 J_{\beta_0^{(r)}} A^- J_{\beta_0^{(r)}}^T),$$

其中 $A = E\{J_{\beta_0^{(r)}}^{\mathrm{T}}[X - E(X|\beta_0^{\mathrm{T}}X)][X - E(X|\beta_0^{\mathrm{T}}X)]^{\mathrm{T}}J_{\beta_0^{(r)}}g'(\beta_0^{\mathrm{T}}X)^2w(\beta_0^{\mathrm{T}}X)\}$, A^- 表示矩阵 A 的广义逆.

需要指出的是: 定理 4.5.2 给出的 $\hat{\beta}$ 的渐近方差与定理 3.10.2 的渐近方差一样, 且与协变量 Z 无关. 定理 3.10.3 已经证明了 $\hat{\beta}$ 比 Härdle 等 (1993) 与 Carroll 等 (1997) 给出的估计量更有效, 即它比现有估计有更小的渐近方差.

由定理 4.5.2, 我们得到 $\hat{\beta}$ 和 β_0 之间夹角的渐近结果, 它可以用来研究充分降维问题. 有关更详细的论述可参见 Cook(1998, 2007).

推论 4.5.1 假设定理 4.5.2 的条件成立. 则

$$|\hat{\beta}^{\mathrm{T}}\beta_0| - 1 = O_P(n^{-1/2}),$$

其中 $|\hat{\beta}^{\mathrm{T}}\beta_0|$ 是 $\hat{\beta}$ 与 β_0 的绝对内积.

下面两个定理给出了 $g(\cdot)$ 的估计量 $\hat{g}^*(\cdot)$ 的收敛速度和 σ^2 的估计量的渐近正态性.

定理 4.5.3 假设定理 4.5.1 的条件成立. 如果 $||\hat{\beta} - \beta_0|| = O_P(n^{-1/2})$, 则

$$\sup_{u\in\mathcal{U}_w, \beta\in\mathcal{B}_n} |\hat{g}(u;\beta) - g(u)| = O_P((nh/\log n)^{-1/2}),$$

其中 $\mathcal{B}_n = \{\beta : ||\beta - \beta_0|| \leqslant cn^{-1/2}\}$, $c > 0$ 为某个常数.

定理 4.5.4 假设条件 (C4.5.1)~(C4.5.6) 成立, 且 $0 < \mathrm{Var}(\varepsilon^2) < \infty$, 则

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2)/(\mathrm{Var}(\varepsilon^2))^{1/2} \xrightarrow{D} N(0,1).$$

可以证明 $n^{-1}\widetilde{\boldsymbol{Z}}^{\mathrm{T}}\widetilde{\boldsymbol{Z}} \xrightarrow{P} \Sigma$. 由定理 4.5.1 和定理 4.5.4 可得

$$(\widetilde{\boldsymbol{Z}}^{\mathrm{T}}\widetilde{\boldsymbol{Z}})^{1/2}(\hat{\theta} - \theta_0)/\hat{\sigma} \xrightarrow{D} N(0, I_q).$$

再利用 (Arnold, 1981) 的定理 10.2d 可得

$$(\hat{\theta} - \theta_0)^{\mathrm{T}}(\widetilde{\boldsymbol{Z}}^{\mathrm{T}}\widetilde{\boldsymbol{Z}})(\hat{\theta} - \theta_0)/\hat{\sigma}^2 \xrightarrow{D} \chi_q^2,$$

其中 χ_q^2 表示自由度为 q 的 χ^2 变量. 用 $\chi_q^2(1-\alpha)(0 < \alpha < 1)$ 表示 χ_q^2 的 $1-\alpha$ 分位数, 那么, θ_0 的近似 $1-\alpha$ 置信域可定义为

$$\{\theta : (\hat{\theta} - \theta)^{\mathrm{T}}(\widetilde{\boldsymbol{Z}}^{\mathrm{T}}\widetilde{\boldsymbol{Z}})(\hat{\theta} - \theta)/\hat{\sigma}^2 \leqslant \chi_q^2(1-\alpha)\}.$$

为构造 β_0 的置信域, 需要 $\hat{\beta}$ 渐近方差的嵌入估计. A 的估计量定义为

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n} J_{\hat{\beta}^{(r)}}^{\mathrm{T}} \widetilde{X}_i \widetilde{X}_i^{\mathrm{T}} J_{\hat{\beta}^{(r)}} \hat{g}'(\hat{\beta}^{\mathrm{T}}X_i; \hat{\beta}, \hat{\theta})^2 w(\hat{\beta}^{\mathrm{T}}X_i),$$

其中 $\hat{g}'(\cdot;\hat{\beta},\hat{\theta})$ 在 (4.5.2) 式中定义, $\widetilde{X}_i = X_i - \hat{\mu}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})$, $\hat{\mu}(u;\hat{\beta})$ 在 (4.6.1) 式中定义, $J_{\hat{\beta}^{(r)}}$ 是 $J_{\beta_0^{(r)}}$ 的估计量. 容易证明 $J_{\hat{\beta}^{(r)}} \xrightarrow{P} J_{\beta_0^{(r)}}$, $\hat{A} \xrightarrow{P} A$. 对任何 $p \times l (l < p)$ 满秩矩阵 D, 由定理 4.5.2 和定理 4.5.4 可以推出

$$(n^{-1}D^{\mathrm{T}}J_{\hat{\beta}^{(r)}}\hat{A}^{-}J_{\hat{\beta}^{(r)}}^{\mathrm{T}}D)^{-1/2}D^{\mathrm{T}}(\hat{\beta} - \beta_0)/\hat{\sigma} \xrightarrow{D} N(0, I_l).$$

再利用 (Arnold, 1981) 的定理 10.2d 可得

$$(\hat{\beta} - \beta_0)^{\mathrm{T}}D(n^{-1}D^{\mathrm{T}}J_{\hat{\beta}^{(r)}}\hat{A}^{-}J_{\hat{\beta}^{(r)}}^{\mathrm{T}}D)^{-1}D^{\mathrm{T}}(\hat{\beta} - \beta_0)/\hat{\sigma}^2 \xrightarrow{D} \chi_l^2.$$

上述结果可以用来构造 $D^{\mathrm{T}}\beta_0$ 的渐近置信域.

4.5.3 模拟研究

考虑部分线性单指标模型

$$Y = 60\exp\{-(\beta_0^{\mathrm{T}}X - 0.5)^2\} + \theta_0 Z_i + \varepsilon,$$

其中 $\beta_0 = (0.75, 0.5, -0.25, -0.25, 0.25)^{\mathrm{T}}$, $\theta_0 = 1$, X 服从 $[-1,1]^5$ 上均匀分布, ε 服从正态分布 $N(0, 0.6^2)$, $g(u) = 60\exp\{-(u-0.5)^2\}$.

我们使用两个方法来对 β_0 和 θ_0 进行估计：4.5 节提出的估计方程估计 (EE) 和 4.4 节提出的最小平均方差估计 (MAVE) 方法. 取权函数 $w(u) = 1$. 对于估计方程估计方法, 通过定义 Z 为二值变量来规定 X 与 Z 之间的依赖关系, 即 Z 取值 $\exp(X^{\mathrm{T}}\beta_Z)/(1+\exp(X^{\mathrm{T}}\beta_Z))$ 的概率为 1, 否则为 0. 考虑 β_Z 与 β_0 之间关系的两个极端情况, 一是 β_Z 与 β_0 平行, 此时选取 $\beta_Z = \beta_0$; 二是 β_Z 与 β_0 正交, 此时选取 $\beta_Z = (0.5, 0, 0.5, 0.5, -0.5)^{\mathrm{T}}$. 对于 β_Z 与 β_0 既不平行也不正交的情况, 模拟结果分析与上述两种情况基本一致.

我们利用局部线性方法对非参数函数做光滑, 核函数取为 $K(u) = 0.75(1-u^2)_+$, 用广义交叉核实 (GCV) 方法选择带宽 h. 例如, 当计算 $g(\cdot)$ 的估计值时, GCV 统计量定义为

$$\mathrm{GCV}(h) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - \hat{\theta}^{\mathrm{T}}Z_i - \hat{g}_h(\hat{\beta}^{\mathrm{T}}X_i; \hat{\beta}, \hat{\theta})]^2/[n^{-1}\mathrm{tr}(I - S_h)]^2, \tag{4.5.9}$$

其中 $\hat{g}_h(\cdot; \hat{\beta}, \hat{\theta})$ 是 $g(\cdot)$ 的具有带宽 h 的估计, 其定义如 (4.5.7) 式, S_h 是相应于带宽 h 的光滑矩阵. 对于 MAVE 以及 EE 估计中的 $\hat{g}$ 和 $\hat{\theta}$, 我们使用最优带宽 h_{opt}. 当计算 EE 估计 $\hat{\beta}$ 时, 选择带宽

$$h = h_{\mathrm{opt}} n^{1/5} n^{-1/3} = h_{\mathrm{opt}} n^{-2/15}, \quad h_1 = h_{\mathrm{opt}}. \tag{4.5.10}$$

这样选取的带宽既满足了定理的条件, 又有好的模拟效果. 相关的讨论可参见 Carroll 等 (1997) 和 Wang 等 (2010).

我们用绝对内积 $|\hat{\beta}^{\mathrm{T}}\beta_0|$ 的均值 (Mean) 和标准差 (SD) 来度量估计量 $\hat{\beta}$ 的精度. 对于估计量 $\hat{\theta}$, 计算其偏差 (Bias)、标准差 (SD) 和均方误差 (MSE). 分别在样本量 $n = 50, 80, 150$ 下进行 500 次运行, 其计算结果放在表 4.5.1 和表 4.5.2 中.

从表 4.5.1 可以看出, MAVE 估计比 EE 估计有更大的内积均值和更小的标准差. 两种估计的内积均值都接近于 1, 其标准差都较小, 并随着 n 的增大而减小. 这说明 β_0 的估计量与真值的内积精度较高. 此外, β_Z 与 β_0 正交比平行具有更大的内积均值和更小的标准差. 表 4.5.2 表明：θ_0 的 MAVE 估计比 EE 估计有更小的 SD 和 MSE, 但在多数情况下 MANE 估计的偏差比 EE 估计的偏差大. β_Z 与 β_0 正交的情形相比于平行, EE 估计量的偏差、SD 和 MSE 更小, 但 MAVE 估计量的偏

表 4.5.1 估计量 $\hat{\beta}$ 和真实值 β_0 的绝对内积的均值 (Mean) 和标准差 (SD)

n	方法	β_Z 与 β_0 平行		β_Z 与 β_0 正交	
		Mean	SD	Mean	SD
50	EE	0.99434	0.06047	0.99525	0.05606
	MAVE	0.99877	0.01318	0.99902	0.01106
80	EE	0.99557	0.05498	0.99723	0.01774
	MAVE	0.99889	0.01255	0.99926	0.01008
150	EE	0.99655	0.03809	0.99886	0.00931
	MAVE	0.99987	0.00016	0.99988	0.00015

表 4.5.2 估计量 $\hat{\theta}$ 的偏差 (Bias)、标准差 (SD) 和均方误差 (MSE)

n	方法	β_Z 与 β_0 平行			β_Z 与 β_0 正交		
		Bias	SD	MSE	Bias	SD	MSE
50	EE	−0.02793	0.60310	0.36451	−0.05098	0.58869	0.34915
	MAVE	0.03936	0.23997	0.05913	0.04178	0.25737	0.06798
80	EE	0.02062	0.56192	0.31618	−0.01777	0.40138	0.16143
	MAVE	0.01903	0.18550	0.03477	0.04137	0.20863	0.04524
150	EE	−0.01572	0.20099	0.04064	−0.00208	0.16636	0.02768
	MAVE	0.01810	0.12283	0.01542	0.03669	0.17750	0.02615

差、SD 和 MSE 更大. 此外, 两种估计量的偏差、SD 和 MSE 都随着样本量的增大而减小.

4.6 经验似然推断

本节利用经验似然方法构造模型 (4.1.1) 中参数分量的经验似然比统计量, 讨论所提出的统计量的渐近性质, 并作模拟研究. 本节的主要素材来自文献 (Zhu and Xue, 2006).

4.6.1 方法论

众所周知, 为利用正态逼近方法构造参数 β_0 和 θ_0 的置信域, 需要构造相应估计量的渐近方差的嵌入估计, 其中包括参数分量和非参数分量的估计, 这就带来一定的麻烦. 经验似然方法避免了这一缺点, 其结构不包含参数 β_0 和 θ_0 的估计. 这是经验似然与正态逼近方法截然不同的地方. 此外, 我们利用 "纠偏" 方法来构造经验似然比使其渐近于标准 χ^2 变量. 下面介绍这个方法.

采用 3.10.1 小节中 "去一分量" 的方法对 β 再参数化, 可得 $\beta^{(r)}$ 的函数 $\beta = \beta(\beta^{(r)})$, 其中 $\beta^{(r)} = (\beta_1, \cdots, \beta_{r-1}, \beta_{r+1}, \cdots, \beta_p)^{\mathrm{T}}$. 引入辅助随机向量

$$\hat{\eta}_i(\beta^{(r)},\theta)=[Y_i-\hat{g}(\beta^{\mathrm{T}}X_i;\beta,\theta)-\theta^{\mathrm{T}}Z_i]\times\begin{pmatrix}J^{\mathrm{T}}_{\beta^{(r)}}[X_i-\hat{\mu}(\beta^{\mathrm{T}}X_i;\beta)]\hat{g}'(\beta^{\mathrm{T}}X_i;\beta,\theta)w(\beta^{\mathrm{T}}X_i)\\ Z_i-\hat{\nu}(\beta^{\mathrm{T}}X_i;\beta)\end{pmatrix},$$

其中 $J_{\beta^{(r)}}$ 在 (3.10.3) 式中定义, $w(\cdot)$ 为具有有界支撑 $\mathcal{U}_w$ 的有界权函数, $\hat{\mu}(t;\beta)$ 和 $\hat{\nu}(t;\beta)$ 分别是 $\mu(u)=E(X|\beta^{\mathrm{T}}X=u)$ 和 $\nu(u)=E(Z|\beta^{\mathrm{T}}X=u)$ 的估计量, 即

$$\hat{\mu}(t;\beta)=\sum_{i=1}^{n}W_{ni}(t;\beta)X_i,\quad \hat{\nu}(t;\beta)=\sum_{i=1}^{n}W_{ni}(t;\beta)Z_i, \tag{4.6.1}$$

其中 $W_{ni}(u;\beta)$ 在 (4.2.3) 式中定义. 因此, 一个纠偏的经验对数似然比函数定义为

$$\hat{l}(\beta^{(r)},\theta)=-2\max\left\{\sum_{i=1}^{n}\log(np_i)\bigg|\,p_i\geqslant 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\hat{\eta}_i(\beta^{(r)},\theta)=0\right\}.$$

记 $V(\beta,\theta)=E\{\varepsilon^2[\Lambda-E(\Lambda|\beta^{\mathrm{T}}X)][\Lambda-E(\Lambda|\beta^{\mathrm{T}}X)]^{\mathrm{T}}\}$, 其中 $\Lambda=(g'(\beta^{\mathrm{T}}X)X^{\mathrm{T}}J_{\beta^{(r)}},Z^{\mathrm{T}})^{\mathrm{T}}$. Zhu 和 Xue(2006) 给出了下列结果.

定理 4.6.1　设条件 (C4.5.1)(i) 和 (C4.5.2)~(C4.5.4) 成立; 并设 $nh^3\to\infty$, $nh^8\to 0$, $V(\beta,\theta)$ 是正定矩阵, $\int_{-\infty}^{\infty}|u|^iK(u)\mathrm{d}u<\infty$, $i\geqslant 1$. 如果 β_0 的第 r 个分量是正的, 则

$$\hat{l}(\beta_0^{(r)},\theta_0)\xrightarrow{D}\chi^2_{p+q-1},$$

应用定理 4.6.1, 可以构造 $(\beta_0^{(r)},\theta_0)$ 的近似 $1-\alpha$ 置信域

$$\{(\beta^{(r)},\theta)|\,\hat{l}(\beta^{(r)},\theta)\leqslant\chi^2_{p+q-1}(1-\alpha),||\beta^{(r)}||<1\}.$$

得到 $(\beta_0^{(r)},\theta_0)$ 的置信域后, 再通过关系式 $\beta_r=(1-\|\beta^{(r)}\|^2)^{1/2}$ 可立即得到 (β_0,θ_0) 的置信域. 因为使用 χ^2_{p+q-1} 而不是 χ^2_{p+q} 来确定这个域, 所以 $(\beta_0^{(r)},\theta_0)$ 的维数减少, 从而置信域的精度提高.

为构造 β_0 的置信域, 我们使用 "去一分量" 方法. 显然, 使用极坐标是另一个可选择的办法, β_0 的维数也可以减少到 $p-1$. 因此, 通过相应极坐标的置信域构造 β_0 的置信域. 然而, 兴趣参数是 β_0 自身, 我们需要把这个极坐标域转变成 β_0 的置信域, 这就需要仔细计算.

4.6.2　模拟研究

考虑 "正弦跳跃" 模型

$$Y_i=18\sin(\pi(\beta^{\mathrm{T}}X_i-a)/(b-a))+\theta Z_i+\varepsilon_i,\quad i=1,\cdots,n, \tag{4.6.2}$$

其中 X_i 是二维变量, 各分量相互独立且都服从均匀分布 $U(0,1)$, $Z_i=0$ 或 $Z_i=1$ 视 i 取奇数或偶数而定, ε_i 服从正态分布 $N(0,0.6^2)$, $\beta=(\beta_1,\beta_2)^{\mathrm{T}}=(\sqrt{3}/3,\sqrt{6}/3)^{\mathrm{T}}$, $\theta=350$. 我们取 $a=\sqrt{3}/2-1.645/\sqrt{12}$, $b=\sqrt{3}/2+1.645/\sqrt{12}$.

模型 (4.6.2) 被 Carroll 等 (1997) 使用过. 他们考虑了参数的点估计. 这里我们通过模拟给出 $\{\beta,\theta\}$ 的置信域及其覆盖概率, 并比较经验似然和正态逼近的优劣.

在模拟计算中, 取 $n=30,60,120$. 通过 200 和 2000 次运行计算 $\{\beta,\theta\}$ 的置信度为 0.95 和 0.99 的置信域及其覆盖概率, 其中图 4.6.1 的结果是经过 200 次运行, 表 4.6.1 的结果是经过 2000 次运行. 为构造置信域, 我们考虑两个方法: 纠偏的经验似然 (BGEL) 和正态逼近 (NA). 取二阶核函数 $K(u)=(15/16)(1-u^2)^2I(|u|\leqslant 1)$. 使用交叉验证 (CV) 方法选择带宽. 假设 $\hat{h}_{\mathrm{opt}}$ 是由 CV 方法选择的最优带宽, 即

$$\hat{h}_{\mathrm{CV}}=\arg\min_h\left\{\frac{1}{n}\sum_{i=1}^{n}\left[Y_i-\hat{g}_h^{(-i)}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta})-\hat{\theta}^{\mathrm{T}}Z_i\right]^2\right\}$$

其中 $\hat{g}_h^{(-i)}(\cdot;\hat{\beta},\hat{\theta})$ 是 $g(\cdot)$ 的具有带宽 h 的 "去一点" 估计量. 为在欠光滑下执行 $\{\beta,\theta\}$ 的最小二乘估计, 在模拟实验中需要使用带宽 $\hat{h}_{\mathrm{opt}}n^{-2/15}$, 参看 Carroll 等 (1997). 这个选择保证所需要的带宽对最佳渐近执行有一个合适的数量级. 基于 NA 的置信域由 $\{\beta,\theta\}$ 的最小二乘估计量 $\{\hat{\beta},\hat{\theta}\}$ 的渐近分布所确定. $\{\hat{\beta}^{(r)},\hat{\theta}\}$ 的渐近方差为 $\hat{\Sigma}=n^{-1}\hat{B}^{-1}\hat{V}\hat{B}^{-1}$, 其中

$$\hat{B}=\frac{1}{n}\sum_{i=1}^{n}[\hat{\Lambda}_i-\hat{\zeta}(\hat{\beta}^{\mathrm{T}}X_i)][\hat{\Lambda}_i-\hat{\zeta}(\hat{\beta}^{\mathrm{T}}X_i)]^{\mathrm{T}},$$

$$\hat{V}=\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2[\hat{\Lambda}_i-\hat{\zeta}(\hat{\beta}^{\mathrm{T}}X_i)][\hat{\Lambda}_i-\hat{\zeta}(\hat{\beta}^{\mathrm{T}}X_i)]^{\mathrm{T}},$$

$\hat{\varepsilon}_i=Y_i-\hat{g}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta})-\hat{\theta}^{\mathrm{T}}Z_i$, $\hat{\Lambda}_i=\left(\hat{g}'(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta})X_i^{\mathrm{T}}J_{\hat{\beta}^{(r)}},Z_i^{\mathrm{T}}\right)^{\mathrm{T}}$, $\hat{\zeta}(\hat{\beta}^{\mathrm{T}}X_i)=\left(\hat{g}'(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta})\hat{\mu}^{\mathrm{T}}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})J_{\hat{\beta}^{(r)}},\hat{\nu}^{\mathrm{T}}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})\right)^{\mathrm{T}}$, $\hat{g}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta})$, $\hat{g}'(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta})$, $\hat{\mu}^{\mathrm{T}}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})$ 和 $\hat{\nu}^{\mathrm{T}}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})$ 分别在 (4.2.2), (4.5.2) 和 (4.6.1) 式中定义.

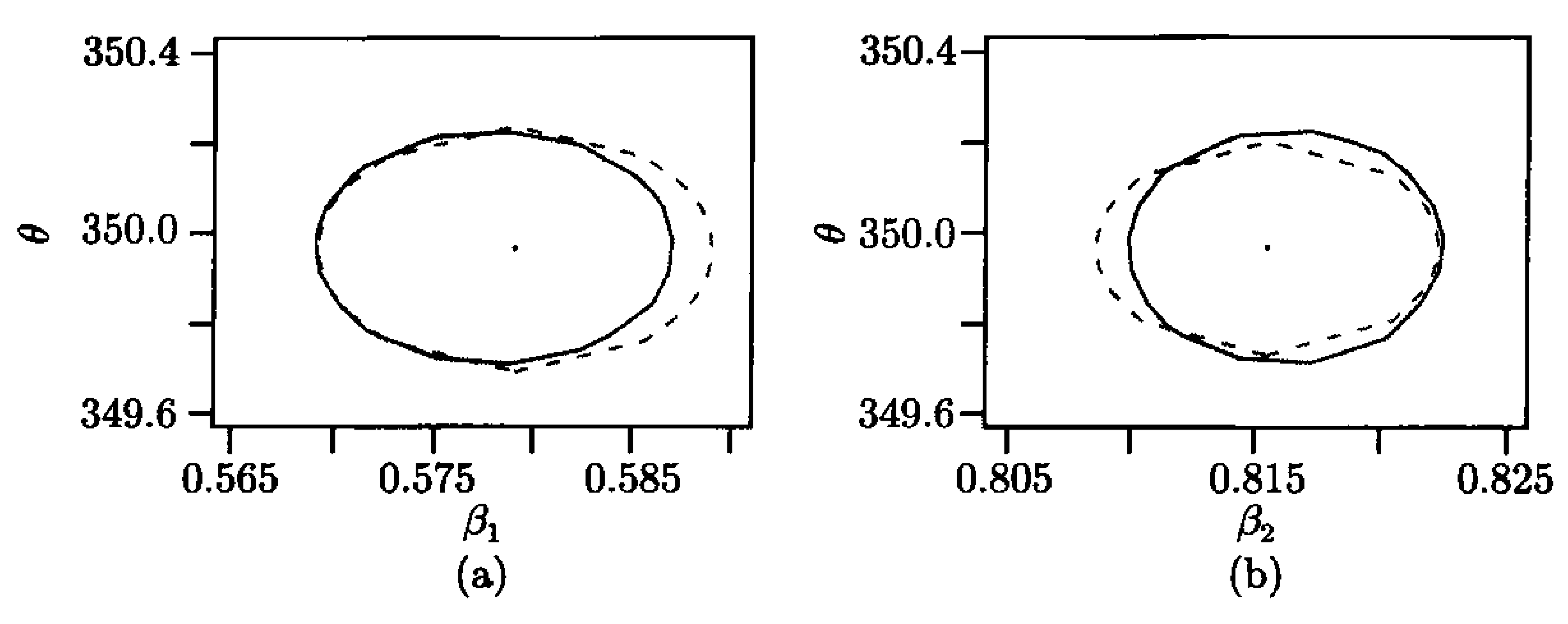


图 4.6.1 基于 BCEL(实线) 和 NA(虚线) 及样本量 $n=120$, 参数 (β_1,θ) 和 (β_2,θ) 的置信域

图 4.6.1 展示了 $\{\beta,\theta\}$ 的置信域, 其中 $n = 120$. 数值结果在表 4.6.1 中报告.

表 4.6.1 参数 (β_1,θ) 和 (β_2,θ) 的近似 0.95 置信域的覆盖概率

n	(β_1,θ)		(β_2,θ)	
	BCEL	NA	BCEL	NA
30	0.9200	0.8960	0.9200	0.8995
60	0.9287	0.9162	0.9287	0.9162
120	0.9330	0.9230	0.9330	0.9230

观察由两种方法得到的 0.95 置信域及其覆盖概率, 我们看到 BCEL 比 NA 给出更小的置信域且达到高的覆盖概率. 当 n 增大时, 覆盖概率增大. 我们的有限模拟研究表明 BCEL 胜过 NA.

4.7 部分线性单指标 EV 模型

本节研究模型 (4.1.1) 中协变量 Z_i 具有测量误差下参数分量和非参数分量的估计问题. 当 Z_i 具有测量误差时, 人们可观察到它的替代变量 W_i. 考虑部分线性单指标 EV 模型:

$$\begin{cases} Y_i = \theta_0^{\mathrm{T}} Z_i + g(\beta_0^{\mathrm{T}} X_i) + \varepsilon_i, \\ W_i = Z_i + U_i, \quad i = 1, \cdots, n, \end{cases} \tag{4.7.1}$$

其中 U_i 为对称的随机误差, 具有协方差阵 Σ_{uu}, 且与 $\{(X_i, Z_i, Y_i), 1 \leqslant i \leqslant n\}$ 独立. 本节假定 Σ_{uu} 是已知的. 对 Σ_{uu} 未知的情况, 可以用它的估计量代替它, 从而可类似地讨论. 详细内容参见 (Liang et al., 1999).

4.7.1 Pseudo-θ 方法

注意到 $E(\varepsilon_i|X_i) = 0$, (4.7.1) 的第一式两边求关于 X_i 的条件期望, 可得

$$E(Y_i|X_i) = \theta_0^{\mathrm{T}} E(Z_i|X_i) + g(\beta_0^{\mathrm{T}} X_i).$$

将 (4.7.1) 的第一式两边分别减去上式两边, 可得

$$Y_i - E(Y_i|X_i) = \theta_0^{\mathrm{T}} [Z_i - E(Z_i|X_i)] + \varepsilon_i. \tag{4.7.2}$$

此时 (4.7.2) 是一个线性模型, 其中 $E(Y|X)$ 和 $E(Z|X)]$ 为未知的回归函数. 当 Z_i 被完全观察到时, 可以用非参数光滑方法构造这两个未知函数的估计, 并记作 $\hat{E}(Y|X)$ 和 $\hat{E}(Z|X)]$. 因此, θ_0 的估计量定义为

$$\hat{\theta}^* = \left\{\sum_{i=1}^{n}[Z_i - \hat{E}(Z_i|X_i)][Z_i - \hat{E}(Z_i|X_i)]^{\mathrm{T}}\right\}^{-1}$$
$$\times \sum_{i=1}^{n}[Z_i - \hat{E}(Z_i|X_i)][Y_i - \hat{E}(Y_i|X_i)]. \tag{4.7.3}$$

因为 Z_i 具有测量误差, 所以如果忽略了测量误差并用 W_i 代替 Z_i, 所得到的估计量不是 θ_0 的相合估计. 因为 (4.7.2) 式对于部分线性模型 ($p=1$) 和部分线性单指标模型都成立. 因此 Liang 等 (1999) 在部分线性模型中提出的纠偏方法仍可以在这里使用. 注意到 (4.7.3) 式与 (Liang et al., 1999) 的 (3) 式类似, 由此可定义 θ_0 的估计量

$$\hat{\theta}_P = \left\{\sum_{i=1}^{n}[W_i - \hat{E}(W_i|X_i)][W_i - \hat{E}(W_i|X_i)]^{\mathrm{T}} - n\Sigma_{uu}\right\}^{-1}$$
$$\times \sum_{i=1}^{n}[W_i - \hat{E}(W_i|X_i)][Y_i - \hat{E}(Y_i|X_i)]. \tag{4.7.4}$$

$E(Y|X=x)$ 的非参数估计具有形式 $\hat{E}(Y|X=x) = \sum_i w_{ni}(x)Y_i$. 通常核权可选择为 $w_{ni}(x) = n^{-1}K_H(X_i - x)$, 其中 $K_H(u) = |H|^{-1/2}K(uH^{-1/2})$, $K(\cdot)$ 是具有紧支撑且有界 Hesse 矩阵的有界 p 元函数, $\int K(u)\mathrm{d}u = 1$, H 是依赖于 n 的 $p\times p$ 对称正定带宽矩阵. 见 (Ruppert and Wand, 1994). 为计算简单起见, 假定 H 是对角矩阵, K_1 是 p 个对称单变量核的乘积.

Liang 和 Wang(2005) 在一些正则条件下证明了 $\hat{\theta}_P$ 的渐近正态性, 即

$$\sqrt{n}(\hat{\theta}_P - \theta_0) \xrightarrow{D} (0, \Gamma^{-1}\Sigma_{\theta P}\Gamma^{-1}), \tag{4.7.5}$$

其中 $\Gamma = E\{[Z - E(Z|X)]^{\otimes 2}\}$, $\Sigma_{\theta P} = E\{(\varepsilon - \theta_0^{\mathrm{T}}U)[Z - E(Z|X)]^{\mathrm{T}}\}^{\otimes 2} + E\{(UU^{\mathrm{T}} - \Sigma_{uu})\theta_0\}^{\otimes 2} + E(UU^{\mathrm{T}}\varepsilon^2)$, $A^{\otimes 2}$ 表示 AA^{T}.

如果 ε 是同方差的且独立于 (X, Z), 则 $\Sigma_{\theta P}$ 可变为 $\sigma_*^2\Gamma + \Sigma_M$, 其中 $\sigma_*^2 = E(\varepsilon - \theta_0^{\mathrm{T}}U)^2$, $\Sigma_M = E\{(UU^{\mathrm{T}} - \Sigma_{uu})\theta_0\}^{\otimes 2} + \Sigma_{uu}\mathrm{var}(\varepsilon)$.

上述结果从理论上说明: 当选择适当的带宽时, 无论 X 的维数怎样, $\hat{\theta}_P$ 都有同样的渐近分布. 然而, 在实际中, 如果 X 的维数较高, 那么就需要很大的样本量. 当 p 较小时, $\hat{\theta}_P$ 提供了一个简单的估计.

得到 θ_0 的估计 $\hat{\theta}_P$ 以后, 就可以使用修正的模型 $Y_i = \hat{\theta}_P^{\mathrm{T}}Z_i + g(\beta_0^{\mathrm{T}}X_i) + \varepsilon_i$ 和 $W_i = Z_i + U_i$ 来估计 β_0 和 $g(\cdot)$. 它实际上是一个单指标模型. 我们就可以应用第 3 章的方法构造 β_0 和 $g(\cdot)$ 的估计量. 这里我们使用局部线性光滑方法估计 $g(\cdot)$, 使用 3.10 节的估计方程方法估计 β_0. 具体估计步骤如下.

步骤 1.0. 使用 $Y_i-\hat{\theta}_P^{\mathrm{T}}Z_i$ 作为响应变量得到 β_0 的一个初始估计 $\hat{\beta}_0$, 如 SIR 估计, 并置 $\hat{\beta}=\hat{\beta}_0/||\hat{\beta}_0||$.

步骤 1.1. 极小化

$$\sum_{i=1}^{n}[Y_i-\hat{\theta}_P^{\mathrm{T}}W_i-a_0-a_1(\hat{\beta}^{\mathrm{T}}X_i-u)]^2K_{1,h_1}(\hat{\beta}^{\mathrm{T}}X_i-u),$$

得到 $\hat{g}(u;\hat{\beta},\hat{\theta}_P)=\hat{a}_0$, $\hat{g}'(u;\hat{\beta},\hat{\theta}_P)=\hat{a}_1$, 其中 $K_{1,h_1}(\cdot)=h_1^{-1}K_1(\cdot/h_1)$, $K_1(\cdot)$ 是一元核函数, h_1 是带宽.

步骤 1.2. 解方程组

$$\begin{cases}\displaystyle\sum_{i=1}^{n}[Y_i-\hat{\theta}_P^{\mathrm{T}}W_i-\hat{g}(\beta^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta}_P)]\hat{g}'(\beta^{\mathrm{T}}X_i;\hat{\beta},\hat{\theta}_P)J_{\hat{\beta}^{(r)}}^{\mathrm{T}}X_iw(\hat{\beta}^{\mathrm{T}}X_i)=0,\\ ||\beta||-1=0,\end{cases}$$

得到 β_0 的更新估计 $\hat{\beta}$, 其中 $J_{\hat{\beta}^{(r)}}$ 在 (3.10.3) 式中定义, $w(\cdot)$ 为具有有界支撑 $\mathcal{U}_w$ 的有界权函数.

步骤 1.3. 重复步骤 1.1 和步骤 1.2 直到收敛.

在上面的迭代程序中, $\hat{\theta}_P$ 是固定的, 这犹如文献中提出的 pseudo 似然估计 (Gong and Samaniego, 1981). 因此称上述估计量为 pseudo-θ 估计量, 并分别把他们记为 $\hat{\beta}_P$, $\hat{\theta}_P$ 和 $\hat{g}_P$.

4.7.2 修正的拟似然方法

在 4.7.1 小节中, 我们首先直接得到 θ_0 的估计, 然后估计 β_0 和 $g(\cdot)$. 虽然这个方法简单且直观, 但 θ_0 的估计没有完全使用给定在模型 (4.1.1) 中的信息, 且依赖于高维非参数估计. 因此期望获得更有效的估计量. 如果 $g(\cdot)$ 是已知的且没有测量误差, 那么 (β_0,θ_0) 的拟似然估计是下列极小化问题的解:

$$Q(\beta,\theta,g,Y,Z,X)=\frac{1}{n}\sum_{i=1}^{n}[Y_i-\theta^{\mathrm{T}}Z_i-g(\beta^{\mathrm{T}}X_i)]^2.$$

当 Z_i 具有测量误差时, 可以用 W_i 代替 Z_i. 注意到 $Q(\beta,\theta,g,Y,Z,W)-\theta^{\mathrm{T}}\Sigma_{uu}\theta$ 与 $Q(\beta,\theta,g,Y,Z,X)$ 有相同的渐近性质. 因此可以用前者代替后者作为目标函数, 并称之为修正的拟似然方法. 下面的迭代算法描述了该方法的估计步骤.

步骤 2.0. 给定初始估计 $(\hat{\beta}_1,\hat{\theta}_L)$, 并置 $\hat{\beta}_L=\hat{\beta}_1/||\hat{\beta}_1||$.

步骤 2.1. 极小化

$$\sum_{i=1}^{n}[Y_i-\hat{\theta}_L^{\mathrm{T}}W_i-a_0-a_1(\hat{\beta}_L^{\mathrm{T}}X_i-u)]^2K_{2,h_2}(\hat{\beta}_L^{\mathrm{T}}X_i-u),$$

得到 $\hat{g}(u;\hat{\beta}_L,\hat{\theta}_L)=\hat{a}_0$, 其中 $K_{2,h_2}(\cdot)=h_2^{-1}K_2(\cdot/h_2)$, $K_2(\cdot)$ 是一元核函数, h_2 是带宽.

步骤 2.2. 极小化

$$\frac{1}{n}\sum_{i=1}^{n}[Y_i-\theta^{\mathrm{T}}W_i-\hat{g}(\beta^{\mathrm{T}}X_i;\hat{\beta}_L,\hat{\theta}_L)]^2-\theta^{\mathrm{T}}\Sigma_{uu}\theta,$$

得到 (β_0,θ_0) 的更新估计 $(\hat{\beta}_L,\hat{\theta}_L)$.

步骤 2.3. 重复步骤 2.1 和步骤 2.2 直到收敛.

在步骤 2.0 中, 要求初始估计 $\hat{\beta}_L$ 和 $\hat{\theta}_L$ 是相合的且具有比 $\sqrt{nh_2}+h_2^2$ 更快的收敛速度, 其中 h_2 是使用在步骤 2.1 中的带宽. 这种选择确保了由步骤 2.1 得到的 $\hat{g}(u;\hat{\beta}_L,\hat{\theta}_L)$ 的渐近性质, 也就是 $\hat{g}(u;\hat{\beta}_L,\hat{\theta}_L)$ 与 $\hat{g}(u;\beta_0,\theta_0)$ 有相同的收敛速度. 在 4.7.1 小节中给出的估计量是 $\sqrt{n}$ 相合的, 且可以在这里充当这样的初始估计的角色.

Liang 和 Wang(2005) 对两种方法得到的估计量的优良性进行了讨论. 得到了如下结论: 当协变量 X 的维数较高时, 估计量 $\hat{\theta}_L$ 比 $\hat{\theta}_P$ 更有效, 并且 $\hat{\theta}_L$ 考虑了降维, 而 $\hat{\theta}_P$ 则出现维数灾祸. 然而, 修正的拟似然方法需要使用迭代算法, 而 pseudo-θ 方法不需要迭代且计算简单. 当 X 和 Z 相互独立时, 这两个估计量在大样本情形下是等价的. 从定理 3.10.3 可知, 估计量 $\hat{\theta}_P$ 比 $\hat{\theta}_L$ 更有效.

4.8 删失数据分析

本节考虑随机删失的部分线性单指标模型, 即在模型 (4.1.1) 中 Y_i 为生存时间. 假定 ε 与 (X,Z) 独立, $E(\varepsilon)=0$, $\mathrm{var}(\varepsilon)=\sigma^2$. 用 C 表示与生存时间 Y 相联系的随机删失时间. 假设 C 与 (X,Z,Y) 独立. 记 $V=\min(Y,C)$, $\Delta=I(Y\leqslant C)$. $\{(X_i,Z_i,V_i,\Delta_i),1\leqslant i\leqslant n\}$ 是来自总体 (X,Z,V,Δ) 的 iid 样本.

因为 Y 随机删失, 所以为构造兴趣参数的估计量, 我们首先进行数据变换. 用 $F(t)$ 和 $G(t)$ 分别表示 Y 和 C 的分布函数, 即 $F(t)=P(Y\leqslant t)$, $G(t)=P(C\leqslant t)$. 记 $\tau_F=\inf\{t:F(t)=1\}$, $\tau_G=\inf\{t:G(t)=1\}$. 本节假设 $\tau_F\leqslant\tau_G$. 当 $G(t)$ 未知时, 我们定义它的 Kaplan-Meier 估计

$$\hat{G}(t)=1-\prod_{i-1}^{n}\left[\frac{n-i}{n-i+1}\right]^{I(V_{(i)}\leqslant t,\Delta_{(i)}=0)}$$

其中 $V_{(1)}\leqslant V_{(2)}\leqslant\cdots\leqslant V_{(n)}$ 是 V 的样本的次序统计量, $\Delta_{(i)}$ 是 Δ 对应于 $V_{(i)}$ 的随机变量, $i=1,\cdots,n$. 因此, 可构造下列伪响应:

$$V_{i\hat{G}}=(1+\phi)L_{i\hat{G}}-\phi K_{i\hat{G}}, \tag{4.8.1}$$

其中 $L_{i\hat{G}} = \int_{-\infty}^{\infty} \{I(V_i \geqslant s)/[1-\hat{G}(s-)] - I(s<0)\}\mathrm{d}s$, $K_{i\hat{G}} = V_i\Delta_i/[1-\hat{G}(V_i-)]$, ϕ 为讨厌参数, 它是用来控制删失或不删失观察的权重. 注意到 $E(Z_{iG}|X_i, Z_i) = E(Y_i|X_i, Z_i)$, 因此当 $G(t)$ 已知时, 上面的数据变换是无偏的. 这种变换由 Fan 和 Gijbels(1994) 引入. 当 $\phi = -1$ 和 $\phi = 0$ 时, 分别得到 Koul 等 (1981) 变换 (简称 KSV 变换)$K_{i\hat{G}}$ 和 Leurgans(1987) 变换 $L_{i\hat{G}}$. Fan 和 Gijbels(1994) 推荐下列 ϕ 的选择:

$$\hat{\phi} = \frac{\int_{-\infty}^{\infty} \{I(V_i \geqslant s)/[1-\hat{G}(s-)] - I(s<0)\}\mathrm{d}s - V_i}{V_i/[1-\hat{G}(V_i-)] - \int_{-\infty}^{\infty} \{I(V_i \geqslant s)/[1-\hat{G}(s-)] - I(s<0)\}\mathrm{d}s},$$

它是最大的 ϕ 使得 $V_{i\hat{G}} \geqslant V_i$.

对给定的 $\{\beta, \theta\}$, 极小化目标函数

$$\sum_{i=1}^{n} [V_{i\hat{G}} - \theta^{\mathrm{T}} Z_i - a_0 - a_1(\beta^{\mathrm{T}} X_i - u)]^2 K_h(\beta^{\mathrm{T}} X_i - u),$$

可得到局部线性估计 $\hat{g}(u; \beta, \theta) = \hat{a}_0$, 其中 $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ 为 $\mathbf{R}$ 上的核函数, $h = h_n$ 为带宽.

然后, 极小化目标函数

$$\frac{1}{n}\sum_{i=1}^{n} [V_{i\hat{G}} - \theta^{\mathrm{T}} Z_i - \hat{g}(\beta^{\mathrm{T}} X_i; \beta, \theta)]^2,$$

可得到 β_0 和 θ_0 的估计量, 并分别记为 $\hat{\beta}$ 和 $\hat{\theta}$. 具体计算时需要利用迭代来完成. 其迭代算法可参见 4.2 节.

记 $\Lambda = (g'(\beta_0^{\mathrm{T}} X) X^{\mathrm{T}}, Z^{\mathrm{T}})^{\mathrm{T}}$, $H = \Lambda - E(\Lambda|\beta_0^{\mathrm{T}} X)$, $\varepsilon_G = Z_G - \theta_0^{\mathrm{T}} Z - g(\beta_0^{\mathrm{T}} X)$, $\Sigma = E(HH^{\mathrm{T}})$, $\Omega = E[(H\varepsilon_G)(H\varepsilon_G)^{\mathrm{T}}]$. $M(s) = \dfrac{E\{[(1+\phi)J(G,s,V) - \phi K_G]I(s<V)\}}{[1-G(s)][1-F(s-)]}$, $J(G,s,V) = I(s<V)\int_s^V [1-G(t-)]^{-1}\mathrm{d}t$, $\Xi = \int_{-\infty}^{\tau_F} M(s)M(s)^{\mathrm{T}}[1-F(s-)]\mathrm{d}G(s)$. 在一些条件下, Lu 和 Cheng(2007) 证明了 $\{\hat{\beta}, \hat{\theta}\}$ 和 $\hat{g}$ 的渐近正态性, 即有下列结果:

$$\sqrt{n}\begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{D} N(0, \Sigma^{-1}(\Omega - \Xi)\Sigma^{-1}),$$

$$\sqrt{nh}[\hat{g}(u; \hat{\beta}, \hat{\theta}) - g(u) - \kappa_2 g''(u)h^2/2] \xrightarrow{D} N(0, \gamma^2(u)),$$

其中 $\gamma^2(u) = \nu_0 \mathrm{var}((Z_G - \theta_0^{\mathrm{T}} Z)|\beta_0^{\mathrm{T}} X = u)$, $\kappa_2 = \int u^2 K(u)\mathrm{d}u$, $\nu_0 = \int K^2(u)\mathrm{d}u$.

注 4.8.1 当没有删失时, 即 $1-G(\cdot)\equiv 1$, 那么 $\Xi\equiv 0$, 此时 $\{\hat{\beta},\hat{\theta}\}$ 的渐近方差为 $\sigma^2\Sigma^{-1}$. 该结果与 (4.2.6) 一致.

4.9 假设检验

应用 4.2 节描述的估计方法, 本节提出两个假设检验. 一是对回归参数的一般假设检验, 二是非参数函数的检验.

4.9.1 检验参数分量

考虑一般的线性检验

$$H_0: A\zeta=\delta \;\leftrightarrow\; H_1: A\zeta\neq\delta, \tag{4.9.1}$$

其中 $\zeta=(\beta^{\mathrm{T}},\theta^{\mathrm{T}})^{\mathrm{T}}$, A 是已知的 $m\times(p+q)$ 满秩矩阵, δ 是 $m\times 1$ 向量. (4.9.1) 的一个简单例子是检验 β 和 θ 的一些元素是否为 0; 即

$$H_0: \beta_{i_1}=\cdots=\beta_{i_k}=0 \text{ 和 } \theta_{j_1}=\cdots=\theta_{j_s}=0,$$

$$H_1: \text{不是所有}\beta_{i_1},\cdots,\beta_{i_k}\text{和}\theta_{j_1},\cdots,\theta_{j_s}\text{等于 } 0.$$

在 H_0 和 H_1 下, 设 $\zeta_0=(\beta_0^{\mathrm{T}},\theta_0^{\mathrm{T}})^{\mathrm{T}}$ 和 $\zeta_1=(\beta_1^{\mathrm{T}},\theta_1^{\mathrm{T}})^{\mathrm{T}}$ 是相应的参数向量, Ω_0 和 Ω_1 分别为 ζ_0 和 ζ_1 的参数空间. 值得注意的是这里有一个记号的稍微滥用, 因为 ζ_0 在前面表示参数 ζ 的真值. 进一步定义

$$Q(H_0)=\inf_{\Omega_0}Q(\zeta_0)=\sum_{i=1}^{n}[Y_i-\hat{\theta}_0^{\mathrm{T}}Z_i-\hat{g}(\hat{\beta}_0^{\mathrm{T}}X_i;\hat{\zeta}_0)]^2,$$

$$Q(H_1)=\inf_{\Omega_1}Q(\zeta_1)=\sum_{i=1}^{n}[Y_i-\hat{\theta}_1^{\mathrm{T}}Z_i-\hat{g}(\hat{\beta}_1^{\mathrm{T}}X_i;\hat{\zeta}_1)]^2,$$

其中 $\{\hat{\beta}_0,\hat{\theta}_0\}$ 和 $\{\hat{\beta}_1,\hat{\theta}_1\}$ 分别是 $\{\beta_0,\theta_0\}$ 和 $\{\beta_1,\theta_1\}$ 的 profile 最小二乘估计, $\hat{g}(\cdot)$ 是由 (4.2.2) 式定义的 $g(\cdot)$ 的局部线性估计. 因此, 构造检验统计量

$$T_1=\frac{n[Q(H_0)-Q(H_1)]}{Q(H_1)}.$$

Liang 等 (2010) 证明了下列结果: 在一些正则条件下, 有

(a) 在 H_0 下, $T_1\xrightarrow{D}\chi_m^2$;

(b) 在 H_1 下, T_1 以分布收敛于自由度为 m 的非中心 χ^2 分布, 其非中心参数

$$\phi=\lim_{n\to\infty}n\sigma^{-2}(A\zeta-\delta)^{\mathrm{T}}(AD^{-1}A^{T})^{-1}(A\zeta-\delta),$$

其中 $D=E\left[\left(g'(\beta_0^{\mathrm{T}}X)\widetilde{X}_0^{\mathrm{T}},\widetilde{Z}_0^{\mathrm{T}}\right)^{\mathrm{T}}\right]^{\otimes 2}$, $\widetilde{X}_0=X-E(X|\beta_0^{\mathrm{T}}X)$, $\widetilde{Z}_0=Z-E(Z|\beta_0^{\mathrm{T}}X)$.

类似地, 可以构造 Wald 检验统计量

$$W_n=(A\hat{\zeta}-\delta)^{\mathrm{T}}(A\hat{D}^{-1}A^{\mathrm{T}})^{-1}(A\hat{\zeta}-\delta),$$

其中 $\hat{D}$ 为 D 的嵌入估计量. 可以证明: W_n 和 T_1 有同样的渐近分布.

4.9.2 检验非参数分量

未知函数 $g(\cdot)$ 的非参数估计为我们进行数据分析提供了描述和图示的信息. 使用这样的信息, 可能构想出一个呈现初步分析特征的参数模型. 为此目的, 我们引入一个拟合优度检验来核定提出的参数模型的适当性. 不失一般性, 我们在零假设下考虑简单的线性模型. 因此, 零假设和对立假设给定为

$$H_0: g(u)=\gamma_0+\gamma_1 u \ \leftrightarrow \ H_1: g(u)\neq\gamma_0+\gamma_1 u, \text{ 对某个}u, \tag{4.9.2}$$

其中 γ_0 和 γ_1 是未知常数参数.

在 H_1 下, 设 $\{\hat{\beta},\hat{\theta}\}$ 和 $\hat{g}(u)$ 是 $\{\beta,\theta\}$ 和 $g(u)$ 相应的 profile 最小二乘和非参数估计量. 在 H_0 下, 我们使用如在 H_1 下得到的同样的参数估计 $\hat{\beta}$ 和 $\hat{\theta}$, 而 $g(u)$ 的估计量是 $\tilde{g}(u)=\hat{\gamma}_0+\hat{\gamma}_1 u.$, 其中 $\hat{\gamma}_0$ 和 $\hat{\gamma}_0$ 分别是 γ_0 和 γ_1 通过拟合 $Y_i-\hat{\theta}^{\mathrm{T}}Z_i$ 对 $\hat{\beta}^{\mathrm{T}}X_I$ 的普通的最小二乘估计. 那么, 在零假设和对立假设下导出的残差平方和为

$$\mathrm{RSS}(H_0)=\sum_{i=1}^{n}[Y_i-\hat{\theta}^{\mathrm{T}}Z_i-\tilde{g}(\hat{\beta}^{\mathrm{T}}X_i)]^2,$$

$$\mathrm{RSS}(H_1)=\sum_{i=1}^{n}[Y_i-\hat{\theta}^{\mathrm{T}}Z_i-\hat{g}(\hat{\beta}^{\mathrm{T}}X_i)]^2.$$

为检验零假设, 考虑下列广义 F 检验,

$$T_2=\frac{r_K}{2}\frac{[\mathrm{RSS}(H_0)-\mathrm{RSS}(H_1)]}{\mathrm{RSS}(H_1)},$$

其中 $r_K=\left[K(0)-0.5\int K^2(u)\mathrm{d}u\right]\left\{\int[K(u)-0.5K*K(u)]\mathrm{d}u\right\}^{-1}$, $K*K$ 表示 K 的卷积, K 是核函数.

Liang 等 (2010) 证明了下列结果: 在一些正则条件下, 如果 H_0 成立, 则 $T_2\xrightarrow{D}\chi_k^2$, 其中 $k=r_K|\mathcal{U}|\left[K(0)-0.5\int K^2(u)\mathrm{d}u\right]\Big/h$, h 是带宽, $|\mathcal{U}|$ 表示 $\mathcal{U}$ 的长度, $\mathcal{U}$ 是 β^TX 的支撑集.

上述结果揭示了部分线性单指标模型的 Wilks 现象. 进一步, 当用多元回归模型代替零假设下的简单线性模型时, 可以得到类似的结果.

4.10 变量选择

在实际问题中, 真模型常常预先是未知的, 一个欠拟合的模型会产生有偏差的估计和预测值, 而一个过拟合的模型会降低参数估计和预测的效率. 这就启发我们利用惩罚方法来同时估计参数和选择重要变量. 为此, 考虑惩罚 profile 最小二乘函数

$$\mathcal{L}_P(\beta,\theta)=\frac{1}{2}Q(\beta,\theta)+n\sum_{j=1}^{p}p_{\lambda_{1j}}(|\beta_j|)+n\sum_{k=1}^{q}p_{\lambda_{2k}}(|\theta_k|), \tag{4.10.1}$$

其中 $Q(\beta,\theta)$ 在 (4.2.5) 式中定义, $p_\lambda(\cdot)$ 是具有调整参数 λ 的惩罚函数. 下文允许 β 和 θ 的不同分量具有不同调整参数的惩罚函数. 为了仅选择变量 X, 我们简单置 $p_{\lambda_{2k}}(\cdot)=0$, 此时, 惩罚 profile 最小二乘函数变为

$$\mathcal{L}_P(\beta,\theta)=\frac{1}{2}Q(\beta,\theta)+n\sum_{j=1}^{p}p_{\lambda_{1j}}(|\beta_j|). \tag{4.10.2}$$

同理, 如果仅对选择变量 Z 感兴趣, 那么置 $p_{\lambda_{1j}}(\cdot)=0$, 因此

$$\mathcal{L}_P(\beta,\theta)=\frac{1}{2}Q(\beta,\theta)+n\sum_{k=1}^{q}p_{\lambda_{2k}}(|\theta_k|). \tag{4.10.3}$$

在文献中有各种可利用的惩罚函数. 为得到 Fan 和 Li(2001) 的 oracle 性质, 我们利用他们的 SCAD 惩罚, 参见 (1.5.3) 式. 对给定的调整参数, 关于 $\{\beta,\theta\}$ 极小化 $\mathcal{L}_P(\beta,\theta)$ 即可得到惩罚估计量. Liang 等 (2010) 证明了上述变量选择方法具有 oracle 性质.

一个重要的问题是选择调整参数. 可以使用 Wang 等 (2007) 的 BIC 选择法选择调整参数 λ_{1j} 和 λ_{2k}. 由于对具有 $p+q$ 个参数的 BIC 函数极小化会带来计算上的困难. 因此, 这里采用 Fan 和 Li(2004) 的方法, 设 $\lambda_{1j}=\lambda\text{SE}(\hat{\beta}_j^u)$, $\lambda_{2k}=\lambda\text{SE}(\hat{\theta}_k^u)$, 其中 λ 为调整参数, $\text{SE}(\hat{\beta}_j^u)$ 和 $\text{SE}(\hat{\theta}_k^u)$ 分别为 β_j 和 θ_k 的非惩罚 profile 最小二乘估计的标准误, $j=1,\cdots,p$, $k=1,\cdots,q$. 所得到的估计量记为 $\hat{\beta}_\lambda$ 和 $\hat{\theta}_\lambda$. 通过极小化目标函数

$$\text{BIC}(\lambda)=\log(\text{MSE}(\lambda))+(\log(n)/n)\text{DF}_\lambda$$

来选择 λ, 其中 $\text{MSE}(\lambda)=n^{-1}\sum_{i=1}^{n}[Y_i-\hat{\theta}_\lambda^{\text{T}}Z_i-\hat{g}(\hat{\beta}_\lambda^{\text{T}}X_i;\hat{\beta}_\lambda,\hat{\theta}_\lambda)]^2$, DF_λ 是 $\hat{\beta}_\lambda$ 和 $\hat{\theta}_\lambda$ 的所有非零系数的个数. 在特殊情况下, 可以在有界区间 $[0,\lambda_{\max}]$ 上的格子点集合中极小化 $\text{BIC}(\lambda)$ 来选择 λ, 其中 $\lambda_{\max}/\sqrt{n}\to 0$. 在实际中, $\text{BIC}(\lambda)$ 的图形可以用来确定近似的 $\lambda_{\max}$, 以保证 $\text{BIC}(\lambda)$ 在 λ 的中等范围内达到极小. 然后在 $[0,\lambda_{\max}]$

上取 λ 的格子点, 从中选择调整参数, 这样做能够在 BIC(λ) 的极小化中避免过多的计算.

参考文献

Arnold S F. 1981. The Theory of Linear Models and Multivariate Analysis. New York: John Wiley & Sons

Carroll R J, Fan J, Gijbels I, et al. 1997. Generalized partially linear single-index models. Journal of the American Statistical Association, 92: 477~489

Chen X, Cui H J. 2009. Empirical likelihood for partially linear single-index errors-in-variables model. Communications in Statistics-Theory and Methods, 38(15): 2498~2514

Cook R D. 1998. Regression Graphics: Ideas for Studying Regressions Through Graphics. New York: Wiley

Cook R D. 2007. Fisher lecture: dimension reduction in regression. Statistical Science, 22: 1~26

Fan J, Gijbels J. 1994. Censored regression: local linear approximations and their applications. Journal of the American Statistical Association, 89: 560~570

Fan J, Gijbels I. 1996. Local Polynomial Modeling and its Applications. London: Chapman and Hall

Fan J, Li R. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96: 1348~1360

Fan J, Li R. 2004. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. Journal of the American Statistical Association, 99: 710~723

Gong G, Samaniego F J. 1981. Pseudo maximum likelihood estimation: theory and applications. The Annals of Statistics, 9: 861~869

Härdle W, Hall P, Ichimura H. 1993. Optimal smoothing in single-index models. The Annals of Statistics, 21: 157~178

He W Q, Yi G Y. 2011. A pairwise likelihood method for correlated binary data with/without missing observations under generalized partially linear single-index models. Statistica sinica, 21(1): 207~229

Heckman N. 1986. Spline smoothing in a partly linear model. Journal of the Royal Statistical Society, Series A, 48: 244~248

Hristache M, Juditsky A, Polzehl J, et al. 2001a. Structure adaptive approach for dimension reduction. The Annals of Statistics, 29: 593~627

Hristache M, Juditsky A, Spokoiny V. 2001b. Direct estimation of the single-index coefficients in single-index models. The Annals of Statistics, 29: 1537~1566

Hsing T, Carroll R J. 1992. An asymptotic theory for sliced inverse regression. The Annals of Statistics, 20: 1040~1061

Huang Z S. 2011a. Statistical inferences for partially linear single-index models with error-prone linear covariates. Journal of Statistical Planning and Inference, 141(2): 899~909

Huang Z S. 2011b. Empirical likelihood for a partially linear single-index measurement error model with right-censored data. Communications in Statistics-Theory and Methods, 40(6): 1015~1029

Koul H, Susarla V, Van Ryzin J. 1981. Regression analysis with randomly right-censored data. The Annals of Statistics, 9: 1276~1288

Lai P, Wang Q H. 2011. Partially linear single-index model with missing responses at random. Journal of Statistical Planning and Inference, 141(2): 1047~1058

Leung B, Wong H, Zhang R Q, et al. 2010. Smoothing Spline Estimation for Partially Linear Single-index Models. Communications in Statistics-Simulation and Computation, 39(10): 1953~1961

Leurgans S. 1987. Linear models random censoring and synthetic data. Biometrika, 74: 301~309

Li B, Wen S Q, Zhu L X. 2008. On a projective resampling method for dimension reduction with multivariate responses. Journal of the American Statistical Association, 106: 1177~1186

Li G R, Zhu L X, Xue L G, et al. 2010. Empirical likelihood inference in partially linear single-index models for longitudinal data. Journal of Multivariate Analysis, 101(3): 718~732

Li K C. 1991. Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association, 86: 316~342

Li K C, Aragon Y, Shedden K, et al. 2003. Dimension reduction for multivariate response data. Journal of the American Statistical Association, 98: 99~109

Li T T, Yang H, Wang J L, et al. 2011. Correction on Estimation for a partial-linear single-index model. The Annals of Statistics, 39(6): 3411~3443

Liang H, Härdle W, Carroll R J. 1999. Estimation in a semiparametric partially linear errors-in-variables model. The Annals of Statistics, 27: 1519~1535

Liang H, Liu X, Li R, et al. 2010. Estimation and testing for partially linear single-index models. The Annals of Statistics, 38: 3811~3836

Liang H, Wang N S. 2005. Partially linear single-index measurement error models. Statistica sinica, 151: 99~116

Liu X H, Wang Z H, Hu X M. 2011. Testing serial correlation in partially linear single-index errors-in-variables models. Communications in Statistics-Theory and Methods, 40(14): 2554~2573

Lu X, Cheng T L. 2007. Randomly censored partially linear single-index models. Journal

of Multivariate Analysis, 98(10): 1895~1922

Ruppert D, Wand M P. 1994. Multivariate weighted least squares regression. The Annals of Statistics, 22: 1346~1370

Speckman P. 1988. Kernel smoothing in partial linear models. Journal of the Royal Statistical Society, Series B, 50: 413~434

Tian P, Yang L, Xue L G. 2010. Asymptotic properties of estimators in partially linear single-index model for longitudinal data. Acta Mathematica Scientia, 30(3): 677~687

Wang H S, Li R, Tsai C L. 2007. Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika, 94: 553~568

Wang J L, Xue L G, Zhu L X, et al. 2010. Estimation for a partial-linear single-index model. The Annals of Statistics, 38(1): 246~274

Xia Y C, Härdle W. 2006. Semi-parametric estimation of partially linear single-index models. Journal of Multivariate Analysis, 97: 1162~1184

Xia Y C, Tong H, Li W K. 1999. On extended partially linear single-index models. Biometrika, 86(4): 831~842

Xia Y C, Tong H, Li W K, et al. 2002. An adaptive estimation of dimension reduction space (with discussions). Journal of the Royal Statistical Society, Series B, 64: 1~28

Yin X R, Zhu L X. 2007. Estimating directions in extending generalized partially linear single-index models. Journal of Computational and Graphical Statistics, 16(2): 330~349

Yu Y, Ruppert D. 2002. Penalized spline estimation for partially linear single-index models. Journal of the American Statistical Association, 97: 1042~105

Yu Y, Ruppert D. 2004. Root-n consistency of penalized spline estimator for partially linear single-index models under general euclidean space. Statistica Sinica, 14(2): 449~455

Zhao W H, Zhang R Q, Huang Z S. 2011. Partially linear single-index beta regression model and score test. Journal of Multivariate Analysis, 103(1): 116~123

Zhu L X, Ng K W. 1995. Asymptotics for Sliced Inverse Regression. Statistica Sinica, 5: 727~736

Zhu L X, Xue L G. 2006. Empirical likelihood confidence regions in a partially linear single-index model. Journal of the Royal Statistical Society, Series B, 68: 549~570

Zou Q M, Zhu Z Y, Wang J L. 2009. Local influence analysis for penalized Gaussian likelihood estimation in partially linear single-index models. Annals of the Institute of Statistical Mathematics, 61(4): 905~918

第 5 章　变系数模型

5.1　引　　言

变系数模型 (varying-coefficient model) 最早起源于 Shumway(1988) 的著作, 随后由 Cleveland 等 (1991) 与 Hastie 和 Tibshirani(1993) 率先对该模型进行了研究. 变系数模型具有如下线性结构:

$$Y=\sum_{j=1}^{p} a_j(U)X_j+\varepsilon, \tag{5.1.1}$$

其中 $(U, X_1, \cdots, X_p)$ 是协变量, Y 是响应变量, $a_j(u)$ 是 $\mathbf{R}$ 上的未知可测函数, ε 是随机误差, 且 $E(\varepsilon|U, X_1, \cdots, X_p)=0$ 和 $\operatorname{var}(\varepsilon|U, X_1, \cdots, X_p)=\sigma^2(U)$. 就 $X_1\equiv 1$ 而论, 允许模型 (5.1.1) 有变截距项.

变系数模型之所以有吸引力是因为诸系数 $a_j(U)$ 依赖于 U, 这不但削减了建模偏差, 而且避免了 "维数祸根". 这个模型的另一个优点是它的可解释性. 当人们对探究回归系数如何随诸如年龄的不同群体而变化感兴趣时, 就自然想到这个模型. 它尤其出现在纵向数据 (longitudinal data) 研究中, 允许人们考察协变量随时间影响响应变量的程度. 变系数模型是一类非常广泛的模型, 包含了常见的一些模型, 如可加模型 (Hastie and Tibshirani, 1990)、部分线性模型 (Härdle et al., 2000)、单指标函数系数回归模型 (Xia and Li, 1999)、自适应的变系数线性模型 (Fan et al., 2003) 等.

自从提出变系数模型以来, 许多统计学者在独立数据和纵向数据下对该模型进行了研究, 取得了丰富的研究成果. Fan 和 Zhang(1999) 提出了系数函数的两步估计方法, 构造了系数函数的估计量, 得到了估计量的渐近均方误及其最优收敛速度. Cai 等 (2000b) 研究了非线性时间序列数据下的函数系数回归模型, 提出了模型拟合优度检验的新的 bootstrap 统计量以及基于新定义的期望预测误差的交叉验证估计的带宽选择方法; 同时, 在 α 混合样本下建立了所提出的估计量的渐近性质. Fan 和 Zhang(2000a) 研究了模型 (5.1.1) 中系数函数的置信带构造和假设检验问题. Cai 等 (2000a) 研究了广义变系数模型的估计和检验问题, 使用局部多项式回归技术构造了系数函数的估计, 建立了所给估计量的渐近正态性; 提出了基于非参数极大似然比检验类的拟合优度检验方法来检验模型中系数函数是否为常数, 同时使用条件 bootstrap 方法估计检验的零分布. 随后, Cai(2002) 用两步估计

方法构造了系数函数的估计量, 并得到了估计量的渐近正态性、均方误差和最优收敛速度. Zhang 和 Lee(2000) 针对系数函数的局部多项式估计研究了变带宽选择问题. Huang 和 Shen(2004) 针对非线性时间序列下的函数系数回归模型并基于多项式样条提出了系数函数估计的整体光滑方法, 得到了所给估计量的相合性和收敛速度. 张日权和卢一强 (2004) 对变系数模型作了进一步研究. Ip 等 (2007) 将广义似然比检验方法应用到系数函数的检验, 证明了所给检验统计量服从渐近 χ^2 分布. Wang 和 Xia(2009) 结合局部多项式光滑和 LASSO 提出了一个新的变量选择方法, 新方法可以使非参数估计和变量选择同时完成. Zhao 和 Xue(2011) 在测量误差数据下研究了模型 (5.1.1) 的变量选择问题. 他们结合基函数逼近和收缩估计提出了一种纠偏的变量选择方法, 利用调整参数的适当选择, 建立了变量选择过程的相合性, 并得到了正则估计量的最优收敛速度. 有关变系数模型的其他研究成果可参见文献 (Galindo et al., 2001; Zhang, 2002; Cai, 2003; Gelfand et al., 2003; Honda, 2004; Hogan et al., 2004; Lu and Mao, 2004; James and Silverman, 2005; Kauermann et al., 2005; Sentürk and Müller, 2005; Tang and Wang, 2005a, 2005b; Heim et al., 2007; Kim, 2007; Hu et al., 2009; Leng, 2009; Tang and Cheng, 2009a, 2009b; Wang et al., 2009; Wang and Zhang, 2009; Wong et al., 2009; Noh and Park, 2010; Li and Xue, 2010; Zhang and Peng, 2010; Kuruwita et al., 2011; Li et al., 2011; Lv et al., 2011; Park et al., 2011; Sentuerk and Nguyen, 2011; Wei, 2011; Wei et al., 2011; Homg and Lian, 2012) 等.

Wu 等 (1998) 在纵向数据下研究了时间变系数模型

$$Y(t) = \beta^{\mathrm{T}}(t)X(t) + \varepsilon(t), \tag{5.1.2}$$

其中 $\varepsilon(t)$ 是均值为 0 的随机过程, 其方差为 $\sigma^2(t)$ 且协方差为 $\rho_\varepsilon(t_1, t_2)$, $\varepsilon(t)$ 和 $X(t)$ 相互独立. 他们利用局部最小二乘准则构造了 $\beta(t)$ 的核估计并建立了它的渐近分布, 所得结果可以用来构造 $\beta(t)$ 的渐近逐点和同时的置信域. Hoover 等 (1998) 构造了 $\beta(t)$ 的光滑样条和局部多项式估计量, 并提出了相应光滑参数选择的交叉验证准则, 同时也证明了所提出估计量的渐近性质. Wu 等 (2000) 提出了系数函数的两步核估计方法来修正普通的核估计, 其基本思想是首先中心化协变量, 然后基于局部最小二乘准则来估计系数函数. 他们研究了所给估计量的大样本性质, 并通过模拟研究和实际数据分析说明了两步核方法优于普通的最小二乘核方法. Fan 和 Zhang(2000b) 建议用两步估计方法来估计系数函数, 提出了系数函数估计量的标准差的估计方法, 并建立了局部多项式估计的渐近结果. Lin 和 Ying(2001) 对 $\beta(t)$ 的累积函数提出了一个逼近形式的估计, 并证明了所给估计的 $\sqrt{n}$ 相合性和渐近正态性. Huang 等 (2002) 利用基函数逼近方法发展了系数函数估计的整体光滑方法, 提出了基于 bootstrap 的推断方法, 并用来构造 $\beta(t)$ 的置信域和实施假设检验, 同

时也给出了估计量的渐近偏差和方差及其相合性. Huang 等 (2004) 采用多项式样条逼近每一个系数函数, 然后利用最小二乘法来做估计, 证明了所给估计量的相合性、渐近正态性及其收敛速度, 并用所得结果构造了 $\beta(t)$ 的渐近置信区间和置信带. Eubank 等 (2004) 进一步考虑了模型 (5.1.2) 中系数函数的光滑样条估计, 研究了系数函数分量的 Bayesian 置信区间, 获得了曲线估计、拟合值和后验方差的有效计算方法以及选择光滑参数的数据驱动方法. Wu 和 Liang(2004) 考虑了一种具有时间相依光滑协变量的随机变系数模型, 提出了用后移算法来估计系数函数的思想. Xue 和 Zhu(2007) 提出了两种纠偏的经验似然比, 即均值校正的经验似然和残差调整的经验似然, 使得所构造的经验对数似然比渐近于卡方分布, 该结果可以用来构造函数系数分量的逐点置信区间和同时置信带. 有关纵向数据下变系数模型的其他研究成果可参见文献 (Kauermann, 2000; Liang et al., 2003; Guo, 2004; Zhang, 2004; Lee and Shaddick, 2007; Krafty et al., 2008; Sentürk and Müller, 2008; Wang et al., 2008; Xu and Zhu, 2008; Lu and Zhang, 2009; Eggermont et al., 2010; Su and Hogan, 2010) 等.

在许多情况下, 诸如流行病研究, 协变量 X 不依赖于 t, 仅响应变量 Y 随时间 t 做重复测量; 即有协变量独立于时间的变系数模型: $Y(t)=\beta^{\mathrm{T}}(t)X+\varepsilon(t)$. 如果假定 $E(XX^{\mathrm{T}})$ 是可逆的, 那么不难得到 $\beta(t)=\{E(XX^{\mathrm{T}})\}^{-1}E[XY(t)]$. $E(XX^{\mathrm{T}})$ 的一个显然的估计是它的相应样本平均. 因此, 仅通过构造 $E[XY(t)]$ 的光滑估计就可以给出 $\beta(t)$ 的计算简单的估计. Wu 和 Chiang(2000) 基于逐分量最小二乘准则提出了时间变系数的两个核估计, 利用均方误差和积分均方误差研究了所提出的估计量的理论特性, 并构造了系数函数的逐点置信区间. Chiang 等 (2001) 提出了逐分量光滑样条方法来估计系数函数; 他们建议使用 "抛去一个体" 交叉验证方法选择光滑参数, 研究了所构造的估计量的渐近正态性, 并给出了样条估计的风险的渐近表示.

5.2 截面数据分析

设 $\{(U_i,X_i,Y_i),1\leqslant i\leqslant n\}$ 是来自模型 (5.1.1) 的 iid 样本, 则有

$$Y_i=\sum_{j=1}^{p}a_j(U_i)X_{ij}+\varepsilon_i,\quad i=1,\cdots,n, \tag{5.2.1}$$

其中 $X_i=(X_{i1},\cdots,X_{ip})^{\mathrm{T}}$. 本节主要介绍模型 (5.2.1) 中兴趣参数的估计方法、模型拟合优度检验及其相应的结果.

5.2.1 局部线性回归估计

1. 估计方法

下面用局部线性回归方法来估计模型 (5.2.1) 中的系数函数 $a_j(u)$, $j=1,\cdots,p$. 对于给定的点 u_0, 在 u_0 的一个邻域内用线性函数

$$a_j(u) \approx a_j(u_0) + a_j'(u_0)(u-u_0) \equiv a_j + b_j(u-u_0), \quad j=1,\cdots,p$$

局部地逼近 $a_j(u)$. 这就得到下列目标函数:

$$L(\boldsymbol{a},\boldsymbol{b}) = \sum_{i=1}^{n}\left\{Y_i - \sum_{j=1}^{p}[a_j + b_j(U_i-u_0)]X_{ij}\right\}^2 K\left(\frac{U_i-u_0}{h}\right),$$

其中 $K(\cdot)$ 是核函数, $h=h_n>0$ 是带宽序列, $\boldsymbol{a}=(a_1,\cdots,a_p)^{\mathrm{T}}$, $\boldsymbol{b}=(b_1,\cdots,b_p)^{\mathrm{T}}$. 注意到 a_j 和 b_j 依赖于 u_0, 从而 $L(\boldsymbol{a},\boldsymbol{b})$ 也依赖于 u_0. 极小化 $L(\boldsymbol{a},\boldsymbol{b})$ 可以得到估计量 $\hat{\boldsymbol{a}}(u_0)$ 和 $\hat{\boldsymbol{b}}(u_0)$, 其中 $\hat{a}(u_0)$ 的分量给出了 $a_1(u_0),\cdots,a_p(u_0)$ 的估计. 为简化记号, 记 $\beta(u_0)=(\boldsymbol{a}^{\mathrm{T}},\boldsymbol{b}^{\mathrm{T}})^{\mathrm{T}}$, $\hat{\beta}(u_0)=(\hat{\boldsymbol{a}}^{\mathrm{T}}(u_0),\hat{\boldsymbol{b}}^{\mathrm{T}}(u_0))^{\mathrm{T}}$. 由最小二乘理论可得

$$\hat{\beta}(u_0) = (\widetilde{X}^{\mathrm{T}}W\widetilde{X})^{-1}\widetilde{X}^{\mathrm{T}}W\boldsymbol{Y}, \tag{5.2.2}$$

其中 $\widetilde{X}$ 为 $n\times 2p$ 矩阵, 其第 i 行元素为 $(X_i^{\mathrm{T}}, X_i^{\mathrm{T}}(U_i-u_0))$, $\boldsymbol{Y}=(Y_1,\cdots,Y_n)^{\mathrm{T}}$, $W=\mathrm{diag}[K_h(U_1-u_0),\cdots,K_h(U_n-u_0)]$. 因此, $a_j(u_0)$ 的估计量为

$$\hat{a}_j(u_0) = e_{j,2p}^{\mathrm{T}}\hat{\beta}(u_0), \quad j=1,\cdots,p,$$

其中 $e_{j,2p}$ 是第 j 个位置为 1 的 $2p\times 1$ 单位向量.

2. 带宽选择

各种现有的带宽选择技术可以应用于上面的估计. 这里使用 Cai 等 (2000b) 提出的一个简单快捷的带宽选择方法, 它可以看作改良的交叉验证准则, 且适用于平稳时间序列数据. 该方法的基本思想如下: 设 m 和 Q 是两个正整数, 且 $n>mQ$. 通过极小化平均均方误差 (AMS)

$$\mathrm{AMS}(h) = \frac{1}{Q}\sum_{q=1}^{Q}\mathrm{AMS}_q(h)$$

来选择 h, 其中

$$\mathrm{AMS}_q(h) = \frac{1}{m}\sum_{i=n-qm+1}^{n-qm+m}\left\{Y_i - \sum_{j=1}^{p}\hat{a}_{j,q}(U_i)X_{ij}\right\}^2,$$

$\{\hat{a}_{j,q}(\cdot)\}$ 是利用样本 $\{(U_i, X_{i1}, \cdots, X_{ip}, Y_i), 1 \leqslant i \leqslant n-qm\}$ 来计算, 其带宽等于 $h[n/(n-qm)]^{1/5}$. 这种选择带宽的优点是对于不同的样本量按照 h 的最优速度来选择带宽, 即 $h \propto n^{-1/5}$. 在实际操作中, 可以使用 $m=[0.1n]$ 和 $Q=4$. 由于带宽的选取不太依赖 m 和 Q 的选择, 因此简单地取 $m=[0.1n]$ 而不取 $m=1$, 这是为了计算上的方便.

3. 主要结果

下面给出 $\hat{\beta}(u_0)$ 的渐近分布. 用 $f_U(u)$ 表示 U 的密度函数. 记 $\mu_j = \int u^j K(u)\mathrm{d}u$, $\nu_j = \int u^j K^2(u)\mathrm{d}u$, $H = \mathrm{diag}(1, h)\otimes I_p$, 这里 $\otimes$ 表示 Kronecker 乘积. 由记 $\sigma^2(u,x) = \mathrm{var}(Y|U=u, X=x)$,

$$\Omega(u) = E(XX^{\mathrm{T}}|U=u), \quad \Gamma(u) = E(XX^{\mathrm{T}}\sigma^2(U,X)|U=u).$$

定理5.2.1 设 $E(Y^4|U=u, X=x)$ 在 $u=u_0$ 的邻域内有界; $E(||X||^3|U=u)$ 和 $a_j''(u)(j=1,\cdots,p)$ 在 u_0 连续; 函数 $f_U(u)$, $\Omega(u)$, $\Gamma(u)$ 和 $\sigma^2(u,x)$ 在 $u=u_0$ 有连续的一阶导数, 且 $f_U(u_0)>0$ 及 $\Omega(u_0)>0$; 核函数 $K(\cdot)$ 是有界的密度且具有紧支撑; $h\to 0$ 且 $nh\to\infty$, 则

$$\sqrt{nh}\left\{H[\hat{\beta}(u_0)-\beta(u_0)] - \frac{h^2}{2(\mu_2-\mu_1^2)}\begin{pmatrix}(\mu_2^2-\mu_1\mu_3)\mathrm{a}''(u_0)\\(\mu_3-\mu_1\mu_2)\mathrm{a}''(u_0)\end{pmatrix} + o_p(h^2)\right\}$$
$$\xrightarrow{D} N(0, \Delta^{-1}\Lambda\Delta^{-1}),$$

其中

$$\Delta = f_U(u_0)\begin{pmatrix}1 & \mu_1\\ \mu_1 & \mu_2\end{pmatrix}\otimes\Omega(u_0), \quad \Lambda = f_U(u_0)\begin{pmatrix}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{pmatrix}\otimes\Gamma(u_0).$$

进一步地, 如果核函数 $K(\cdot)$ 是对称的, 那么

$$\sqrt{nh}\left\{\hat{\mathrm{a}}(u_0) - \mathrm{a}(u_0) - \frac{h^2\mu_2}{2}\mathrm{a}''(u_0) + o_p(h^2)\right\} \xrightarrow{D} N(0, \Sigma(u_0)),$$

其中

$$\Sigma(u_0) = \nu_0\Omega^{-1}(u_0)\Gamma(u_0)\Omega^{-1}(u_0)/f_U(u_0).$$

当 $K(\cdot)$ 对称时, 估计量 $\hat{a}_j(u_0)$ 的均方误差 (MSE) 是

$$\mathrm{MSE} = \frac{1}{4}h^4\mu_2^2[a_j''(u_0)]^2 + \frac{\nu_0 e_{j,p}^{\mathrm{T}}\Omega^{-1}(u_0)\Gamma(u_0)\Omega^{-1}(u_0)e_{j,p}}{nhf_U(u_0)},$$

其中 $e_{j,p}$ 为 $2p\times 1$ 单位向量, 其第 j 个元素是 1. 由此可得到最优带宽:

$$h_{j,\text{opt}}=\left\{\frac{\nu_0 e_{j,p}^{\mathrm{T}}\Omega^{-1}(u_0)\Gamma(u_0)\Omega^{-1}(u_0)e_{j,p}}{\mu_2^2 f_U(u_0)[a_j''(u_0)]^2}\right\}^{1/5} n^{-1/5}.$$

如果使用最优带宽 $h_{j,\text{opt}}$, 那么 MSE 具有阶 $n^{-4/5}$.

5.2.2 光滑样条估计

系数函数光滑的另一个方法是由 Hastie 和 Tibshirani(1993) 提出的光滑样条法, 即极小化

$$\sum_{i=1}^{n}\left[Y_i-\sum_{j=1}^{p}a_j(U_i)X_{ij}\right]^2+\sum_{j=1}^{p}\lambda_j\int[a_j''(u)]^2\mathrm{d}u, \tag{5.2.3}$$

来获得 $a_1(u),\cdots,a_p(u)$ 的估计, 其中 $\lambda=(\lambda_1,\cdots,\lambda_p)^{\mathrm{T}}$ 是光滑参数. 这是一个颇具权威的思想, 但有几个潜在的问题. 首先, 有 p 个参数需要同时选择, 这在具体实施中有相当大的难度. 其次, 计算上也是一个挑战. Hastie 和 Tibshirani(1993) 提出了一个迭代算法, 并详细描述了计算的细节. 第三, 得到估计量的样本性质有一定困难. 用这个方法得到的结果能否达到与一步方法同样的最优收敛速度, 还是一个有待解决的问题.

5.2.3 多项式样条估计

1. 估计方法

所谓多项式样条, 就是在内节点集合上将多项式与光滑结合在一起的分片多项式. 为了精确表述这个概念, 假设区间 $\mathcal{U}$ 上的节点序列为 $\xi_0<\xi_1<\cdots<\xi_{M+1}$, 其中 ξ_0 和 ξ_{M+1} 是 $\mathcal{U}$ 的端点. $\mathcal{U}$ 上的 $l\geqslant 0$ 次多项式样条就是每个区间 $[\xi_m,\xi_{m+1})$ 和 $[\xi_M,\xi_{M+1}]$ 上的 l 次多项式, 且都有 $l-1(l\geqslant 1)$ 阶连续导数, 此处 $0\leqslant m\leqslant M-1$. 分片常数函数、线性样条、二次样条和三次样条分别对应于 $l=0,1,2,3$. 具有特殊幂次和节点的样条函数的采集来自线性函数空间, 它容易构造合适的基. 例如, 三次样条和节点序列为 $\xi_0,\cdots,\xi_{M+1}$ 的空间来自 $M+4$ 维线性空间. 这个空间的截断权基是 $1,x,x^2,x^3,(x-\xi_1)_+^3,\cdots,(x-\xi_M)_+^3$. 具有较好数字特性的基是 B 样条基. 参见 Boor(1978) 和 Schumaker(1981) 关于样条函数的综合论述.

假设模型 (5.1.1) 中的系数函数 $a_j(u)(j=1,\cdots,p)$ 是光滑的, 则它能很好地用一个样条函数 $a_j^*(u)$ 来逼近, 即当样条的节点数目趋于无穷大时, $\sup\limits_{u\in\mathcal{U}}|a_j^*(u)-a_j(u)|\to 0$(Boor, 1978; Schumaker, 1981). 因此, 存在一个基函数集合 $B_{js}(\cdot)$(即 B 样条) 和常数 $\gamma_{js}^*(s=1,\cdots,K_j)$, 使得

$$a_j(u) \approx a_j^*(u) = \sum_{s=1}^{K_j} \gamma_{js}^* B_{js}(u). \tag{5.2.4}$$

那么, 极小化关于 γ 的函数

$$l(\gamma) = \sum_{i=1}^{n} \left[Y_i - \sum_{j=1}^{p} \left(\sum_{s=1}^{K_j} \gamma_{js} B_{js}(U_i) \right) X_{ij} \right]^2, \tag{5.2.5}$$

可得到 γ 的估计, 其中 $\gamma = (\gamma_1^{\mathrm{T}}, \cdots \gamma_p^{\mathrm{T}})^{\mathrm{T}}$, $\gamma_j = (\gamma_{j1}, \cdots, \gamma_{jK_j})^{\mathrm{T}}$. 假设 (5.2.5) 存在唯一的极小值点, 且记它的极小值点为 $\hat{\gamma} = (\hat{\gamma}_1^{\mathrm{T}}, \cdots \hat{\gamma}_p^{\mathrm{T}})^{\mathrm{T}}$, $\hat{\gamma}_j = (\hat{\gamma}_{j1}, \cdots, \hat{\gamma}_{jK_j})^{\mathrm{T}}$, $j = 1, \cdots, p$. 则 $a_j(u)$ 的估计量可表示为

$$\hat{a}_j(u) = \sum_{s=1}^{K_j} \hat{\gamma}_{js} B_{js}(u), \quad j = 1, \cdots, p.$$

称 $\hat{a}_j(u)$ 为 $a_j(u)$ 的最小二乘样条估计. 关于 $\hat{\gamma}$ 的明确表达式可以类似于 (5.3.16) 式的推导方法而得到.

基表示的思想可以更一般地应用到函数逼近的其他基系, 譬如, 多项式基和 Fourier 基. 这里考虑 B 样条是因为它有优良的样条逼近性质和好的 B 样条基的数字特性. 当使用 B 样条时, (5.2.4) 式中项 K_j 的数目依赖于节点的数目和 B 样条的阶. 注意到不同的 $a_j(u)$ 允许不同的 K_j. 当不同的 $a_j(u)$ 有不同的光滑度时, B 样条就提供了适应性.

2. 节点数的选择

节点数作为光滑参数, 起着与局部线性方法中的带宽一样的作用. 虽然一个主观的光滑参数可以通过检查估计曲线或残差图来确定, 但利用数据选择 K_j 的自动程序也是有实际意义的. 通常是极小化一个准则函数来选择节点数. 这里考虑四个准则函数: AIC(Akaike, 1974), $\mathrm{AIC_C}$(Hurvich and Tsai, 1989), BIC(Schwarz, 1978) 和修正的交叉验证 (MCV)(Cai, et al., 2000b). 用 n 表示 (5.2.5) 式右边项的个数, $k = \sum_j K_j$ 是待估参数的个数, RSS=$l(\hat{\gamma})$ 是 (5.2.5) 式中残差平方和的最小值. 前三个准则定义为

$$\mathrm{AIC} = \log(n^{-1}\mathrm{RSS}) + 2kn^{-1},$$

$$\mathrm{AIC_C} = \mathrm{AIC} + \frac{2(k+1)(k+2)}{n(n-k-2)},$$

$$\mathrm{BIC} = \log(n^{-1}\mathrm{RSS}) + kn^{-1}\log n.$$

MCV 准则可以看作一个修正的多块交叉验证准则, 对时间序列数据非常有用. 设 m 和 Q 是两个给定的正整数且 $n > mQ$. 使用长度为 $n - qm (q = 1, \cdots, Q)$ 的子

序列来估计系数函数 a_j, 基于估计的模型来计算长度为 m 的下一段时间序列的一步预测误差. 那么, MCV 准则函数是 AMS=$\sum_{q=1}^{Q}$AMS$_q$, 其中

$$\text{AMS}_q = \frac{1}{m}\sum_{i=n-qm+1}^{n-qm+m}\left[Y_i - \sum_{j=1}^{p}\left(\sum_{s=1}^{K_j}\hat{\gamma}_{js}^{(q)}B_{js}(U_i)\right)X_{ij}\right]^2,$$

$q=1,\cdots,Q$, $\{\hat{\gamma}_{js}^{(q)}\}$ 是用样本 $\{(U_i,X_i,Y_i),1\leqslant i\leqslant n-qm\}$ 计算得到估计. Cai 等 (2000b) 建议使用 $m=[0.1n]$ 与 $Q=4$.

3. 相合性和收敛速度

为了表达清楚起见, 下面用函数空间记号表示样条估计. 设 $\mathcal{A}_j$ 是 $\mathcal{T}$ 上有固定阶和节点的多项式样条空间, 具有有界的网状比 (即, 连续节点之间的不同的比是有界的且大于 0). 则可由下式获得样条估计 $\hat{a}_j$:

$$\hat{a}_j = \arg\min_{a_j\in\mathcal{A}_j}\sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{p}a_j(U_i)X_{ij}\right)^2 I(U_i\in\mathcal{C}),\quad j=1,\cdots,p.$$

上式本质上与 (5.2.5) 式相同, 但它用到了函数空间记号 (假定 $\{B_{js},1\leqslant s\leqslant K_j\}$ 是 $\mathcal{A}_j$ 的一个基). 这里在最小二乘准则中用权函数是为了屏蔽观察以外的数据, 这是遵循非参数时间序列的常见用法 (Tjϕstheim and Auestad, 1994).

下面在紧区间 $\mathcal{C}$ 上考察样条估计的性质. 设 $||a||_2=\left[\int_{\mathcal{C}}a^2(t)dt\right]^{1/2}$ 是 $\mathcal{C}$ 上平方可积函数的 L_2 模. 如果 $||\hat{a}_j-a_j||_2\xrightarrow{P}0$ $(n\to\infty)$, 则称 $\hat{a}_j$ 是 a_j 的相合估计. 记 $K_n=\max\limits_{1\leqslant j\leqslant p}K_j$, $\rho_{n,j}=\sup\limits_{g\in\mathcal{G}_j}||g-a_j||_2$, $\rho_n=\max\limits_{1\leqslant j\leqslant p}\rho_{n,j}=\inf\limits_{a\in\mathcal{A}_j}||a-a_j||_2$.

定理5.2.2 设① U_i 的密度函数在 $\mathcal{C}$ 上一致有界且大于 0; ② 矩阵 $E(X_iX_i^{\mathrm{T}}|U_i=u)$ 的特征根在 $\mathcal{C}$ 上一致有界且大于 0; ③ $K_n=c_1n^r$, $0<r<1$, $c_1>0$ 为常数; ④ $\{(U_i,X_i,Y_i),1\leqslant i\leqslant n\}$ 是 α 混合过程, 且混合系数 $\alpha(\cdot)$ 满足 $\alpha(n)\leqslant c_2n^{-\delta}$, $\delta>(5/2)r/(1-r)$, $c_2>0$ 为常数; ⑤ 对某个充分大的 $m>0$, $E(|X_{ij}|^m)<\infty$, $j=1,\cdots,p$; (vi) ε_i 与 (U_i,X_i) $(j=1,\cdots,p,i'\leqslant i)$ 和 $\varepsilon_{i'}(i'<i)$ 独立, $E(\varepsilon_i)=0$, $\text{var}(\varepsilon_i)\leqslant c_3$, 这里 $c_3>0$ 为某个常数 (噪声误差可以是异方差), 则

$$||\hat{a}_j-a_j||_2^2=O_P\left(K_nn^{-1}+\rho_n^2\right),\quad j=1,\cdots,p.$$

特别地, 如果 $\rho_n=o(1)$, 那么 $\hat{a}_j$ 是 a_j 的相合估计, 即 $||\hat{a}_j-a_j||_2=o_P(1)$, $j=1,\cdots,p$.

上述定理给出了与 iid 数据同样的收敛速度 (Stone, 1982). 这里 K_n 是用来度量估计空间 $\mathcal{G}_j$ 的大小. ρ_n 是用来度量逼近误差的大小, 它是由诸 a_j 的光滑度和样条空间 $\mathcal{A}_j$ 的维数来确定. 例如, 如果 $a_j(j=1,\cdots,p)$ 有有界连续导数, 那么 $\rho_n=O(K_n^{-2})$ (DeVore and Lorentz, 1993, 第 7 章定理 7.2). 在这种情况下, $\hat{a}_j$ 收敛到 a_j 的速度是 $n^{-1}K_n+K_n^{-4}$. 特别地, 如果 K_n 随着样本量 n 的增加而增加且与 $n^{1/5}$ 同阶, 则 $\hat{a}_j$ 的收敛速度是 $n^{-4/5}$. 定理 5.2.2 的证明可参见 (Huang and Shen, 2004).

5.2.4 拟合优度检验

为检验模型 (5.2.1) 是否具有特殊的参数形式, Cai 等 (2000b) 基于参数和非参数拟合的残差平方和 (RSS) 的比较提出了一个拟合优度检验. 下面介绍这个方法. 考虑零假设

$$H_0: a_j(u)=\varphi_j(u,\theta), \quad 1\leqslant j\leqslant p, \tag{5.2.6}$$

其中 $\varphi_j(u,\theta)$ 是具有未知参数 θ 的给定的函数族. 设 $\hat{\theta}$ 为 θ 的估计量. 在零假设下 RSS 为

$$\mathrm{RSS}_0=\frac{1}{n}\sum_{i=1}^{n}[Y_i-\varphi_1(U_i,\hat{\theta})X_{i1}-\cdots-\varphi_p(U_i,\hat{\theta})X_{ip}]^2.$$

类似地, 相应于模型 (5.2.1) 的 RSS 为

$$\mathrm{RSS}_1=\frac{1}{n}\sum_{i=1}^{n}[Y_i-\hat{a}_1(U_i)X_{i1}-\cdots-\hat{a}_p(U_i)X_{ip}]^2.$$

检验统计量定义为

$$T_n=\frac{\mathrm{RSS}_0-\mathrm{RSS}_1}{\mathrm{RSS}_1}=\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1}-1.$$

对大的 T_n 值拒绝零假设 (5.2.6). 使用下列非参数 bootstrap 方法估计检验的 p 值.

(1) 从中心化残差 $\{\hat{\varepsilon}_i-\bar{\hat{\varepsilon}}_i, 1\leqslant i\leqslant n\}$ 的经验分布中产生 bootstrap 残差 $\{\varepsilon_i^*, 1\leqslant i\leqslant n\}$, 其中

$$\hat{\varepsilon}_i=Y_i-\hat{a}_1(U_i)X_{i1}-\cdots-\hat{a}_p(U_i)X_{ip}, \quad \bar{\hat{\varepsilon}}_i=\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i,$$

并定义

$$Y_i^*=Y_i-\varphi_1(U_i,\hat{\theta})X_{i1}-\cdots-\varphi_p(U_i,\hat{\theta})X_{ip}+\varepsilon_i^*.$$

(2) 基于样本 $\{(U_i,X_i,Y_i^*), 1\leqslant i\leqslant n\}$ 计算 bootstrap 检验统计量 T_n^*.

(3) 当 T_n 大于给定 $\{(U_i,X_i,Y_i), 1\leqslant i\leqslant n\}$ 下 T_n^* 的条件分布的 α 分位点时拒绝零假设 H_0.

检验的 p 值自然是 bootstrap 样本重复下事件 $\{T_n > T_n^*\}$ 的相对频率. 为简单起见, 我们在计算 T_n^* 时使用与 T_n 相同的带宽. 注意到我们使用非参数拟合代替参数拟合的中心化残差, 因为无论零假设或对立假设是否正确, 残差的非参数估计总是相合的. 即使当零假设不成立, 这个方法也提供零分布的相合估计.

5.2.5　模拟研究

考虑变系数模型

$$Y = a_0(U) + a_1(U)X_1 + a_2(U)X_2 + \varepsilon,$$

其中 $U \sim U[0,1]$, X_1 与 X_2 相互独立且都服从标准正态分布, $\varepsilon \sim N(0, 0.8^2)$, $a_0(u) = 8\exp\{-(u-0.5)^2\}$, $a_1(u) = 2\cos(2\pi u)$, $a_2(u) = 5(u-0.5)^2$.

我们采用局部线性光滑和多项式样条来估计系数函数. 在局部线性光滑估计计算中, 取 Epanechnikov 核 $K(u) = 0.75(1-u^2)_+$, 用改良的分块交叉核实 (MCV) 法选取带宽 (见 5.2.1 小节). 对于多项式样条估计, 取三次样条, 用 5.2.3 小节介绍的 MCV 准则选取节点数. 取样本量 $n = 100$, 用 500 次重复计算得到估计曲线 $\{\hat{a}_\nu(\cdot),\ \nu = 0,1,2\}$. 利用根均方误差 (RMSE) 评估 $\{\hat{a}_\nu(\cdot),\ \nu = 0,1,2\}$ 的执行效果,

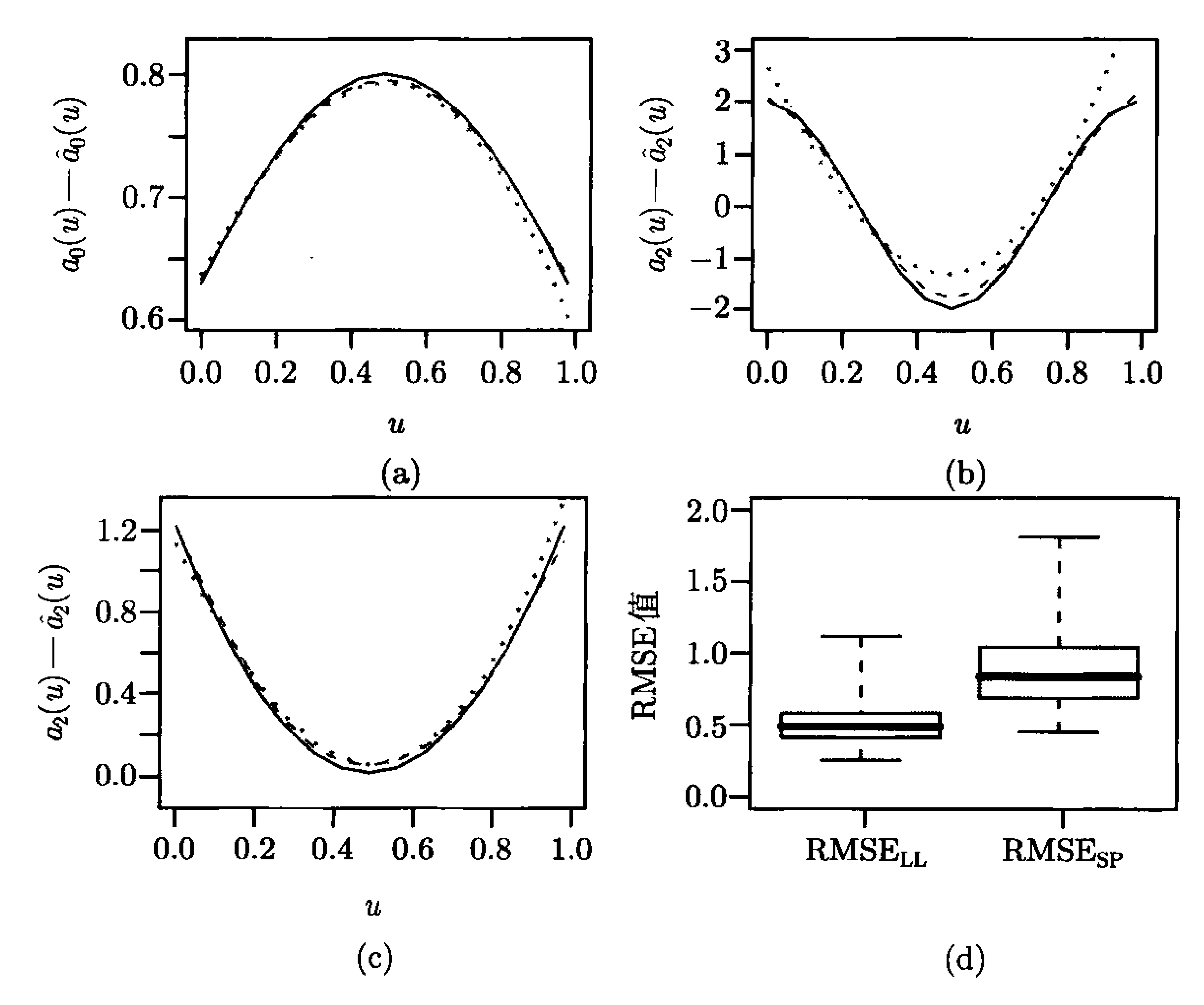


图 5.2.1　当 $n = 100$ 时的模拟结果. (a)~(c): 真实曲线 (实线), 局部线性估计曲线 (虚线), 多项式样条估计曲线 (点线). (a) 对 $a_0(\cdot)$, (b) 对 $a_1(\cdot)$, (c) 对 $a_2(\cdot)$; (d) $\{\hat{a}_\nu(\cdot),\ \nu = 0,1,2\}$ 的 RMSE 箱线图, 左边图形对应局部线性估计方法, 右边图形对应多项式样条估计方法

即

$$\mathrm{RMSE}^2 = n_{\mathrm{grid}}^{-1}\sum_{j=0}^{2}\sum_{k=1}^{n_{\mathrm{grid}}}[\hat{a}_j(u_k) - a_j(u_k)]^2,$$

其中 $\{u_k, k=1,\cdots,n_{\mathrm{grid}}\}$ 为适当的格子点. 模拟结果展示在图 5.2.1 中.

图 5.2.1(a)~(c) 给出了系数函数的真实曲线和估计的曲线. 图 5.2.1(d) 给出了 500 次计算的 RMSE 的箱线图. 从图 5.2.1 可以看出, 系数函数的局部线性估计的 RMSE 的中位数小于多项式样条估计的 RMSE 的中位数, 这表明局部线性估计优于多项式样条估计.

5.3 纵向数据分析

本节利用变系数模型对纵向数据进行分析, 提出模型的系数函数的若干估计量, 其中包括: 局部核估计, 局部多项式估计, 光滑样条估计和最小二乘估计. 本节的主要内容取材于 (Wu et al., 1998; Hoover et al., 1998; Huang et al., 2002) 等.

5.3.1 模型

纵向数据常常在生物医学和计量经济学研究中遇到, 此类数据的例子也可以在临床试验和疾病追踪研究的文献中看到. 考虑来自 n 个个体的数据, 其第 i 个个体具有 m_i 次观测, $1\leqslant i\leqslant n$, 总的观测数为 $N=\sum_{i=1}^{n}m_i$. 设 $t_{ij}(1\leqslant j\leqslant m_i)$ 是第 i 个个体的第 j 次观测时间, $Y_{ij}=Y_i(t_{ij})$ 和 $X_i(t_{ij})$ 分别是第 i 个个体在时间 t_{ij} 的响应变量和协变量的观测, 这里 Y_{ij} 是实值变量, $X_i(t_{ij})$ 是 $p\times 1$ 向量. 虽然由 $\{(t_{ij},X_i(t_{ij}),Y_{ij}),1\leqslant i\leqslant n,1\leqslant j\leqslant m_i\}$ 给出的纵向测量在不同的个体之间是独立的, 但在同一个体内的重复测量可能是相关的. 响应变量和协变量的依赖关系由下面的时间变系数模型给出:

$$Y_{ij}=\beta^{\mathrm{T}}(t_{ij})X_i(t_{ij})+\varepsilon_i(t_{ij}), \tag{5.3.1}$$

其中 $X_i(t)=(1,X_{i1}(t),\cdots,X_{ip}(t))^{\mathrm{T}}$, $X_{il}(t)$ 是时间 t 的实值协变量, $\beta(t)=(\beta_0(t),\cdots,\beta_p(t))^{\mathrm{T}}$ 是未知回归系数向量, 且 $\beta_l(t)\in R$, $l=0,\cdots,p$, 误差 $\varepsilon_i(t)$ 是均值为 0 的随机过程, 且 $\varepsilon_i(t)$ 是独立的. 为一般性, 不需要限定 t_{ij} 是非负的, 即 $t_{ij}\in\mathbf{R}$.

5.3.2 局部核估计

1. *估计方法*

假设 $(X(t),Y(t))$ 与 $(X_i(t),Y_i(t))$ 同分布. 对于每个给定的 $t\in\mathbf{R}$, 模型 (5.3.1) 的等价形式是 (5.1.2). 假设给定 $t\in\mathbf{R}$ 下的条件期望 $E[X(t)X^{\mathrm{T}}(t)]$ 和 $E[X(t)Y(t)]$

存在, 且 $E[X(t)X^{\mathrm{T}}(t)]$ 可逆. 那么, 对于任何 $t \in \mathbf{R}$, $\beta(t)$ 是 $E\{[Y(t)-b^{\mathrm{T}}(t)X(t)]^2\}$ 的唯一极小值点, 且由下式给出:

$$\beta(t) = \{E[X(t)X^{\mathrm{T}}(t)]\}^{-1}E[X(t)Y(t)]. \tag{5.3.2}$$

因此, 估计 $\beta(t)$ 的一个自然方法是利用局部最小二乘准则. 设 $K(\cdot)$ 是一个 Borel 可测的核函数, h 是依赖于 n 和 m_i 的正的带宽. 那么, $\beta(t)$ 的局部核估计可通过极小化

$$L_N(\mathrm{b}(t)) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}[Y_{ij} - b^{\mathrm{T}}(t)X_i(t_{ij})]^2 K\left(\frac{t_{ij}-t}{h}\right) \tag{5.3.3}$$

而得到. 称该估计量为局部核估计, 并记为 $\hat{\beta}_{\mathrm{LK}}(t)$. 可以将 (5.3.3) 式等价地写作如下矩阵形式:

$$L_N(\mathrm{b}(t)) = \sum_{i=1}^{n}[Y_i - \boldsymbol{X}_i b(t)]^{\mathrm{T}}\boldsymbol{K}_i(t)(Y_i - \boldsymbol{X}_i b(t)), \tag{5.3.4}$$

其中 $Y_i = (Y_{i1},\cdots,Y_{im_i})^{\mathrm{T}}$, $\boldsymbol{X}_i$ 是一个 $m_i\times(p+1)$ 矩阵, 其第 j 行元素为 $(1, X_{i1}(t_{ij}), \cdots, X_{ip}(t_{ij}))$, $j=1,\cdots,m_i$, $\boldsymbol{K}_i(\cdot)$ 是一个对角核矩阵, 即

$$\boldsymbol{K}_i(t) = \mathrm{diag}\left\{K((t_{i1}-t)/h),\cdots,K((t_{im_i}-t)/h)\right\}.$$

假定 $\sum\limits_{i=1}^{n}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{K}_i(t)\boldsymbol{X}_i$ 也是可逆的, 那么 $\hat{\beta}_{\mathrm{LK}}(t)$ 作为目标函数 (5.3.4) 的唯一极小值点可由下面 $p+1$ 维列向量给出:

$$\hat{\beta}_{\mathrm{LK}}(t) = \left(\sum_{i=1}^{n}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{K}_i(t)\boldsymbol{X}_i\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{K}_i(t)Y_i\right). \tag{5.3.5}$$

使用 $\hat{\beta}_{\mathrm{LK}}(t)$ 好处是它的数学表达式简单清晰, 在实际中容易实现, 且具有优良的渐近性质. 然而, 因为 $\hat{\beta}_{\mathrm{LK}}(t)$ 仅仅包含一个带宽, 当 $\beta_0(t),\cdots,\beta_p(t)$ 是不同的光滑族时, 它不能对 $\beta(t)$ 的所有分量提供适当的光滑. 因此, 进一步研究其他最小二乘类估计方法是必需的, 譬如光滑样条和局部多项式估计, 它们用多个光滑参数以适应 $\beta_0(t),\cdots,\beta_p(t)$ 的不同光滑的需要.

2. 带宽选择

下面考虑带宽选择问题. 由于个体之间是独立的, 因此由 Rice 和 Silverman (1991) 提出的一个直观的带宽选择方法是 "抛去一个体" 交叉验证 ("leave-one-subject-out" cross-validation). 假设我们要测量 $\hat{\beta}_{\mathrm{LK}}(t)$ 的风险, 常常使用它的平均预测平方误差 (average prediction squared error):

$$\mathrm{APSE}(\hat{\beta}_{\mathrm{LK}}) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}E\{[Y_{ij}^{*} - \hat{\beta}_{\mathrm{LK}}^{\mathrm{T}}(t_{ij})X_i(t_{ij})]^2\},$$

其中 Y_{ij}^* 是在 $(t_{ij}, X_i(t_{ij}))$ 处的一个新观测. 那么, “抛去一个体” 交叉验证准则可由下式给出:

$$\mathrm{CV}(h) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}[Y_{ij} - \hat{\beta}_h^{(-i)}(t_{ij})^{\mathrm{T}}X_i(t_{ij})]^2, \tag{5.3.6}$$

其中 $\hat{\beta}_h^{(-i)}(t)$ 是用删除第 i 个个体以后的所有观测得到的 $\beta(t)$ 的估计, 这里取局部核估计量. 极小化 $\mathrm{CV}(h)$ 可以得到一个交叉验证带宽 h_{CV}, 即 $h_{\mathrm{CV}} = \inf\limits_{h>0}\mathrm{CV}(h)$. 此准则可以很容易地推广到其他光滑估计, 譬如光滑样条和局部多项式.

3. 主要结果

现在给出 $\hat{\beta}_{\mathrm{LK}}(t_0)$ 的渐近分布. 下文假定设计点列 $\{t_{ij}, 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m_i\}$iid, 且具有公共的密度 f. 记 $S(f)$ 为 f 的支撑. 设 t_0 是 $S(f)$ 的内点, 并记

$$\sigma^2(t_0) = E[\varepsilon^2(t_0)], \quad \rho_\varepsilon(t_0) = \lim_{\Delta\to 0} E[\varepsilon(t_0+\Delta)\varepsilon(t_0)],$$

$$\omega_{lr}(t_0) = E[X_{il}(t_{ij})X_{ir}(t_{ij})|t_{ij} = t_0], \quad l, r = 0, \cdots, p,$$

$\Omega(t_0)$ 是一个 $(p+1)\times(p+1)$ 矩阵, 其 (l,r) 元素为 $\omega_{lr}(t_0)$.

为得到主要结果, 下列正则条件是必需的.

(C5.3.1) 对某个常数 $h_0 > 0$, 带宽 $h = h_0 N^{-1/5}$.

(C5.3.2) 对某个 $0 \leqslant \lambda < \infty$, $\lim\limits_{n\to\infty} N^{-6/5}\sum\limits_{i=1}^{n} m_i^2 = \lambda$.

(C5.3.3) 核 $K(\cdot)$ 是有界且对称的概率密度函数, 且满足 $\int u^4 K(u)\mathrm{d}u < \infty$.

(C5.3.4) 存在常数 $\delta \in (2/5, 2]$ 使得 $\sup\limits_t E[|\varepsilon_i(t_{ij})|^{2+\delta}|t_{ij} = t] < \infty$, $\sup\limits_t E[X_{il}^4(t_{ij})|t_{ij} = t] < \infty$, $i = 1, \cdots, n$, $j = 1, \cdots, m_i$, $l = 1, \cdots, p$.

(C5.3.5) 对任何 $l, r = 0, \cdots, p$, $\omega_{lr}(t)$ 和 $f(t)$ 在点 t_0 有连续的一阶导数, 且 $\beta_r(t)$ 在点 t_0 有连续的二阶导数.

(C5.3.6) $\sigma^2(t)$ 和 $\rho_\varepsilon(t)$ 在点 t_0 连续.

(C5.3.7) $\Omega(t_0) = \big(\omega_{lr}(t_0)\big)$ 是 $(p+1)\times(p+1)$ 阶正定矩阵.

定理5.3.1 假设条件 (C5.3.1)~(C5.3.7) 成立, 则

$$\sqrt{Nh}[\hat{\beta}_{\mathrm{LK}}(t_0) - \beta(t_0)] \xrightarrow{D} N(B(t_0), \Omega^{-1}(t_0)\Gamma(t_0)\Omega^{-1}(t_0)),$$

其中 $\Gamma(t_0)$ 是一个 $(p+1)\times(p+1)$ 矩阵, 其 (l,r) 元素为

$$\gamma_{lr}(t_0) = \sigma^2(t_0)\omega_{lr}(t_0)(f(t_0))^{-1}\int K^2(u)\mathrm{d}u + \lambda h_0\rho_\varepsilon(t_0)\omega_{lr}(t_0), \tag{5.3.7}$$

$$B(t_0) = (f(t_0))^{-1}\Omega^{-1}(t_0)(b_0(t_0), \cdots, b_p(t_0))^{\mathrm{T}},$$

$$
\begin{aligned}
b_l(t_0) = h_0^{3/2} \sum_{k=1}^{p} [\beta_k'(t_0)\omega_{lk}'(t_0)f(t_0) + \beta_k'(t_0)\omega_{lk}(t_0)f'(t_0) \\
+(1/2)\beta_k''(t_0)\omega_{lk}(t_0)f(t_0)] \int u^2 K(u)\mathrm{d}u,
\end{aligned}
$$

这里 $l, r = 0, \cdots, p$.

上述结果是由 Wu 等 (1998) 给出的. Wu 等 (1998) 也给出了 $\hat{\beta}_{\rm LK}(t_0)$ 的渐近偏差和方差的估计, 从而可利用定理 5.3.1 构造 $\beta(t_0)$ 的置信域. 同时, Wu 等 (1998) 也构造了 $\beta(t_0)$ 的 Bonferroni 类置信带.

注5.3.1　在定理 5.3.1 中, 如果将条件 (C5.3.1) 改为 $h = o(N^{-1/5})$, 条件 (C5.3.2)~(C5.3.7) 不变, 则渐近偏差项消失, 即有

$$
\sqrt{Nh}[\hat{\beta}_{\rm LK}(t_0) - \beta(t_0)] \xrightarrow{\mathcal{D}} N(0, \Omega^{-1}(t_0)\Gamma(t_0)\Omega^{-1}(t_0)),
$$

注5.3.2　定理 5.3.1 的一个直接含义是: 为了确保 $\hat{\beta}_{\rm LK}(t_0)$ 有好的渐近性质, 重复测量的数目 $m_1, \cdots, m_n$ 必须比样本量 N 相对小. Hoover 等 (1998) 证明了下列结果: $\hat{\beta}_{\rm LK}(t_0)$ 是 $\beta(t_0)$ 的相合估计当且仅当 $\sum_{i=1}^{n} m_i^2 = o(N^2)$, 亦即等价于 $\max_{1\leqslant i\leqslant n}(m_i/N) = o(1)$. 这里假定稍强的条件 $\sum_{i=1}^{n} m_i^2 = o(N^{6/5})$ 是为了确保 $\hat{\beta}_{\rm LK}(t_0)$ 能达到收敛速度 $N^{-2/5}$.

5.3.3　局部多项式估计

我们使用 Fan(1992) 提出的局部多项式拟合方法估计 $\beta_l(t)$, $l = 0, \cdots, p$. 对给定的点 t, 在 t 的邻域内用下列多项式逼近:

$$
\beta_l(T) \approx \sum_{s=0}^{q} b_{ls}(t)(T-t)^s, \quad l = 0, \cdots, p,
$$

其中 $b_{ls}(t) = \beta_l^{(s)}(t)/s!$, $s = 0, \cdots, q$. 考虑下列局部最小二乘问题:

$$
l(b(t)) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[Y_{ij} - \sum_{l=0}^{p}\sum_{s=0}^{q} b_{ls}(t)(t_{ij}-t)^s X_{il}(t_{ij})\right]^2 K\left(\frac{t_{ij}-t}{h}\right), \qquad (5.3.8)
$$

其中 $b(t) = (b_{01}(t), \cdots, b_{0q}(t), \cdots, b_{p1}(t), \cdots, b_{pq}(t))^{\rm T}$, $K(\cdot)$ 为核函数, $h = h_n$ 为带宽. 极小化 $l(b(t))$ 可得到 $b(t)$ 的估计量, 亦即得到 $\beta(t_0)$ 的估计量 $\hat{\beta}(t_0)$. 下面推导 $\hat{\beta}(t)$ 的解析表达式. 为了方便起见, 记 $Y_i = (Y_{i1}, \cdots, Y_{im_i})^{\rm T}$, $X_{ijl} = X_{il}(t_{ij})$, $V_{ij}(t) = (X_{ij0}, \cdots, X_{ij0}(t_{ij}-t)^q, \cdots, X_{ijp}, \cdots, X_{ijp}(t_{ij}-t)^q)^{\rm T}$, $V_i(t) = (V_{i1}(t), \cdots, V_{im_i}(t))^{\rm T}$, $i = 1, \cdots, n, j = 1, \cdots, m_i$, $W_i(t) = {\rm diag}\{K((t_{i1}-t)/h), \cdots, K((t_{im_i} -$

$t)/h)\}$ 是一个对角权矩阵. 那么 (5.3.8) 式可重写为

$$l(b(t)) = \sum_{i=1}^{n}[Y_i - V_i(t)b(t)]^{\mathrm{T}} W_i(t)[Y_i - V_i(t)b(t)].$$

假定 $\sum_{i=1}^{n} V_i^{\mathrm{T}}(t)W_i(t)V_i(t)$ 是可逆的, 那么最小二乘估计 $\hat{\mathrm{b}}(t)$ 是唯一的且为

$$\hat{b}(t) = \left(\sum_{i=1}^{n} V_i^{\mathrm{T}}(t)W_i(t)V_i(t)\right)^{-1} \left(\sum_{i=1}^{n} V_i^{\mathrm{T}}(t)W_i(t)Y_i\right).$$

因此, 局部多项式回归估计可定义为

$$\hat{\beta}_{\mathrm{LP}}(t) = (I_p \otimes e_{1,q})\hat{b}(t),$$

其中 $\otimes$ 为 Kronecker 乘积, $e_{k,q}$ 表示长度为 $q+1$ 的单位向量, 其第 k 个位置的元素为 1. 由 (5.3.5) 式给出的局部核估计 $\hat{\beta}_{\mathrm{LK}}(t)$ 是上述局部多项式估计 $\hat{\beta}_{\mathrm{LP}}(t)$ 当 $q=0$ 时的特殊情况. 估计量 $\hat{\beta}_{\mathrm{LP}}(t)$ 的最优带宽仍可以按 (5.3.6) 式定义的准则来选取. 需要说明的是: $\hat{\beta}_{\mathrm{LP}}(t)$ 的渐近性质还没有作深入研究, 这是一个公开的问题.

5.3.4 光滑样条估计

假设 $\beta_l(t)$ $(l=0,\cdots,p)$ 是二次连续可微的, 其二阶导数 $\beta_l''(t)$ 有界且平方可积. Hoover 等 (1998) 按照 Hastie 和 Tibshirani(1993) 的思想建议通过极小化

$$J(\beta,\lambda) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[Y_{ij} - \sum_{l=0}^{p}\beta_l(t_{ij})X_{ijl}\right]^2 + \sum_{l=0}^{p}\lambda_l\int[\beta_l''(t)]^2\mathrm{d}t, \tag{5.3.9}$$

来获得 $\beta(t) = (\beta_0(t),\cdots,\beta_p(t))^{\mathrm{T}}$ 的估计量, 其中 $\lambda = (\lambda_0,\cdots,\lambda_p)^{\mathrm{T}}$ 是正的光滑参数, 它是用来惩罚 $\beta_0(t),\cdots,\beta_p(t)$ 的光滑程度.

Hastie 和 Tibshirani(1993) 证明了: 在一定的条件下, 由 $J(\beta,\lambda)$ 得到的估计量是唯一的. 称该估计量为光滑样条估计, 并记为 $\hat{\beta}_{\mathrm{SS}}(t)$.

为了极小化 $J(\beta,\lambda)$, 利用样条基函数表示 $\beta_0(t),\cdots,\beta_p(t)$ 是方便的. 用下列形式来表达每一个 $\beta_l(t)$:

$$\beta_l(t) = \sum_{r=1}^{d}\gamma_{lr}B_r(t) = B^{\mathrm{T}}(t)\gamma_l, \quad t\in\mathbf{R}, \tag{5.3.10}$$

这里 $d>1$, $\gamma_l = (\gamma_{l1},\cdots,\gamma_{ld})^{\mathrm{T}}$ 是实值系数, $B(t) = (B_1(t),\cdots,B_d(t))^{\mathrm{T}}$ 是基函数集. 然后极小化二次函数 $J(\beta,\lambda)$ 来找系数向量 γ_l $(l=0,\cdots,p)$. 为此, 记

$$B_i = \begin{pmatrix} B_1(t_{i1}) & \cdots & B_d(t_{i1}) \\ \vdots & & \vdots \\ B_1(t_{im_i}) & \cdots & B_d(t_{im_i}) \end{pmatrix},$$

Ω 是 $d\times d$ 矩阵, 其 (i,j) 元素为 $\Omega_{ij}=\int\{B_i''(t)B_j''(t)\}\mathrm{d}t$. 那么, (5.3.9) 式等价于

$$J(\beta,\lambda)=\sum_{i=1}^{n}\left[Y_i-\sum_{l=0}^{p}\mathbf{X}_{i,l}B_i\gamma_l\right]^{\mathrm{T}}\left[Y_i-\sum_{l=0}^{p}\boldsymbol{X}_{i,l}B_i\gamma_l\right]+\sum_{l=0}^{p}\lambda_l\gamma_l^{\mathrm{T}}\Omega\gamma_l, \quad (5.3.11)$$

其中 $Y_i=(Y_{i1},\cdots,Y_{im_i})^{\mathrm{T}}$, $\boldsymbol{X}_{i,l}=\mathrm{diag}[X_{il}(t_{i1}),\cdots,X_{il}(t_{im_i})]$. 如果令 $\partial J(\beta,\lambda)/\partial\gamma_{rl}=0$, $l=0,\cdots,p$, 那么 (5.3.11) 的极小值点 $(\gamma_0,\cdots,\gamma_p)$ 满足正则方程组

$$\sum_{i=1}^{n}\left[(\boldsymbol{X}_{i,l}B_i)^{\mathrm{T}}\sum_{l=0}^{p}\boldsymbol{X}_{i,l}B_i\gamma_l\right]+\sum_{l=0}^{p}\lambda_l\Omega\gamma_l=\sum_{i=1}^{n}(\boldsymbol{X}_{i,l}B_i)^{\mathrm{T}}Y_i, \quad l=0,\cdots,p. \quad (5.3.12)$$

如果正则方程组 (5.3.12) 有唯一解 $(\hat{\gamma}_0,\cdots,\hat{\gamma}_p)$, 那么, 不失一般性, 存在 $d\times m_i$ 矩阵 N_{il} $(i=1,\cdots,n,l=0,\cdots,p)$, 使得

$$\hat{\gamma}_l=\sum_{i=1}^{n}N_{il}Y_i, \quad l=0,\cdots,p.$$

用 $(\hat{\gamma}_0,\cdots,\hat{\gamma}_p)$ 代替 (5.3.10) 式中的 $(\gamma_0,\cdots,\gamma_p)$, 即可得到相应的估计量, 即

$$\hat{\beta}_l(t)=\sum_{i=1}^{n}B^{\mathrm{T}}(t)N_{il}Y_i, \quad l=0,\cdots,p. \quad (5.3.13)$$

这个线性系统的解 $(\hat{\gamma}_0,\cdots,\hat{\gamma}_p)$ 的存在性和唯一性依赖于设计矩阵 $\boldsymbol{X}_{i,l}$ 和 t_i, $i=1,\cdots,n$. 在光滑样条的实际实施中, 人们必须选择适当的光滑参数向量 λ 和基函数. 从 (5.3.9) 式可以看出: 太大的 λ_l 会对 $\beta_l(t)$ 的凸凹性进行过分的惩罚, 这就导致一个过光滑的估计 $\hat{\beta}_l(t)$. 相反地, 太小的 λ_l 会导致一个欠光滑的估计 $\hat{\beta}_l(t)$.

实际上, 如果 (5.3.12) 的唯一解存在, 那么此解可以直接求得, 也可以使用 Hastie 和 Tibshirani(1990) 提出的后移 (backfitting) 算法得到. 注意到 (5.3.12) 构成 $(p+1)d\times(p+1)d$ 阶方程系, 它可以被用来求出所有估计量的解. 一个实际的问题是当 d 相当大时给解方程组带来了困难.后移算法正是克服这一困难的一个方法. 利用具有相对小的固定同等空间节点的样条来逼近光滑样条解, 也能够妥当地避开 d 太大的困难. 这就充分削减了计算的维数. 关于纵向观测下 $\hat{\beta}_l(t)(l=0,\cdots,p)$ 的深入的理论特性需要进一步研究.

(5.3.6) 式也可以用来选择光滑样条估计的光滑参数. 对于由 (5.3.8) 式得到的局部多项式估计 $\hat{\beta}_{\mathrm{LP}}(t)$, 极小化 $\mathrm{CV}(h)$ 将得到单一的带宽 h_{CV}. 当 $\beta_0(t),\cdots,\beta_p(t)$ 满足不同的光滑度条件时, 相应的估计曲线不能够很好地拟合系数曲线. 对于本节提出的光滑样条估计, 交叉验证光滑参数包含 $\lambda_{0,\mathrm{CV}},\cdots,\lambda_{p,\mathrm{CV}}$. 直观地, 光滑样条中众多的光滑参数能够被用来满足非参数分量的不同光滑度的需要. 从光滑参数上讲, 光滑样条估计优于局部多项式估计.

5.3.5 最小二乘基估计

1. *估计方法*

本节假定 t_{ij} 在一个有限长度的区间 $\mathcal{T}$ 上取值. 假设对每一个 $l=0,\cdots,p$, 有一个基函数 $B_{ls}(t)$ 和常数 γ_{ls}^* 的集合, $s=1,\cdots,K_l$, 使得

$$\beta_l(t) \simeq \sum_{s=1}^{K_l} \gamma_{ls}^* B_{ls}(t), \quad t \in \mathcal{T}.$$

那么, 我们可以用

$$Y_{ij} \simeq \sum_{l=0}^{p}\sum_{s=1}^{K_l} \gamma_{ls}^* B_{ls}(t_{ij}) X_{il}(t_{ij}) + \varepsilon_i(t_{ij})$$

近似模型 (5.3.1), 且通过极小化

$$l(\gamma) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} w_i \left[Y_{ij} - \sum_{l=0}^{p}\sum_{s=1}^{K_l} \gamma_{ls} B_{ls}(t_{ij}) X_{il}(t_{ij})\right]^2 \qquad (5.3.14)$$

来估计 γ_{ls}^*, 其中 $\gamma = (\gamma_0^{\mathrm{T}},\cdots,\gamma_p^{\mathrm{T}})^{\mathrm{T}}$, $\gamma_l = (\gamma_{l1},\cdots,\gamma_{lK_l})^{\mathrm{T}}$, w_i 是第 i 个个体的非负权, 且 $\sum_{i=1}^{n} m_i w_i = 1$. 假设 $l(\gamma)$ 有唯一的极小值点, 并记它的极小值点为 $\hat{\gamma} = (\hat{\gamma}_0^{\mathrm{T}},\cdots,\hat{\gamma}_p^{\mathrm{T}})^{\mathrm{T}}$, 其中 $\hat{\gamma}_l = (\hat{\gamma}_{l1},\cdots,\hat{\gamma}_{lK_l})$, $l=0,\cdots,p$. 那么, 自然用

$$\hat{\beta}_l(t) = \sum_{s=1}^{K_l} \hat{\gamma}_{ls} B_{ls}(t)$$

来估计 $\beta_l(t)$, 并称 $\hat{\beta}_l(t)$ 为 $\beta_l(t)$ 的最小二乘基估计.

为了给出 $\hat{\gamma}$ 的明确表达式, 引入下面一些记号. 记 $W_i = \mathrm{diag}\{w_i,\cdots,w_i\}$, $Y_i = (Y_{i1},\cdots,Y_{im_i})^T$,

$$B(t) = \begin{pmatrix} B_{01}(t) & \cdots & B_{0K_0}(t) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & B_{p1}(t) & \cdots & B_{pK_p}(t) \end{pmatrix}.$$

$$U_i(t_{ij}) = [X_i^{\mathrm{T}}(t_{ij}) B(t_{ij})]^{\mathrm{T}}, \quad U_i = (U_i(t_{i1}),\cdots,U_i(t_{im_i}))^{\mathrm{T}}.$$

那么 (5.3.14) 式等价于

$$l(\gamma) = \sum_{i=1}^{n} (Y_i - U_i\gamma)^T W_i (Y_i - U_i\gamma). \qquad (5.3.15)$$

假定 $\sum_{i=1}^{n} U_i^{\mathrm{T}} W_i U_i$ 是可逆的, 那么最小二乘估计 $\hat{\gamma}$ 是唯一的且定义为

$$\hat{\gamma} = \left(\sum_{i=1}^{n} U_i^{\mathrm{T}} W_i U_i\right)^{-1} \left(\sum_{i=1}^{n} U_i^{\mathrm{T}} W_i Y_i\right). \tag{5.3.16}$$

利用这个矩阵表达式, 可以将 $\beta(t)$ 的最小二乘基估计写为

$$\hat{\beta}_{\mathrm{LSB}}(t) = (\hat{\beta}_0(t), \cdots, \hat{\beta}_p(t))^{\mathrm{T}} = B(t)\hat{\gamma}. \tag{5.3.17}$$

注5.3.3　由基函数 $\{B_{l1}, \cdots, B_{lK_l}\}$ 生成的线性函数空间 $\mathcal{C}_l$ 可唯一地确定基估计 $\hat{\beta}_l(t)(0 \leqslant l \leqslant p)$. 不同的基函数集可以用来生成同一个空间 $\mathcal{C}_l$, 因此, 虽然相应的 $\hat{\gamma}$ 可能不同, 但却给出同一个估计量 $\hat{\beta}_l(t)$. 例如, B 样条基和截断权基都可以被用来生成样条函数空间.

注5.3.4　在 $\hat{\gamma}$ 和 $\hat{\beta}_{\mathrm{LSB}}(t)$ 的理论和实际特性上, 对 (5.3.14) 式中 w_i 的选择需要一个重要的说明. 选择 $w_i \equiv 1/N$ 相当于对每一个观测有相同的权, 而取 $w_i \equiv 1/(nm_i)$ 相当于对每一个个体有相同的权. 可以想象, w_i 的理想选择可能依赖于数据的个体内相关结构. 然而, 真实的相关结构在实际中常常是未知的, 如果 $m_i(i \equiv 1, \cdots, n)$ 相对较小, $w_i \equiv 1/N$ 似乎是一个实用的选择, 否则, 取 $w_i \equiv 1/(nm_i)$ 是适合的.

2. 选择基

可以使用任何一个函数逼近的基系. 当基本函数具有周期性时, Fourier 基是可取的, 多项式基也是常用的选择, 它能对光滑函数提供好的逼近. 然而, 这些基对展示某些局部性质可能不是太敏感, 除非使用大的 K_l. 从这方面讲, 多项式样条是值得提倡的. 的确, 如何用一个相对小的 K_l 来选择基使能达到优良的逼近是一个重要的问题. 对一般的指导性建议, 可参见文献 (Ramsay and Silverman, 1997) 的 §3.2.2.

3. 选择光滑参数

根据 5.3.2 小节提出的思想, 我们使用 "抛去一个体" 交叉验证方法来选择 K_l. 设 $\hat{\gamma}^{(-i)}$ 是由 (5.3.16) 式定义的最小二乘基估计, 其中删除了第 i 个个体的观测. 用 $\hat{\gamma}^{(-i)}$ 代替 (5.3.17) 式中的 $\hat{\gamma}$ 可得到估计量 $\hat{\beta}_{\mathrm{LSB}}^{(-i)}(t)$. 定义

$$\mathrm{CV}(K) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_i \left[Y_{ij} - \hat{\beta}_{\mathrm{LSB}}^{(-i)}(t_{ij})^{\mathrm{T}} X_i(t_{ij})\right]^2 \tag{5.3.18}$$

作为 $K = (K_0, \cdots, K_p)$ 的交叉验证得分. 极小化 $\mathrm{CV}(K)$ 可得到交叉验证光滑参数 K_{CV}. 使用这个交叉验证方法有两个主要理由: 第一, 删除一个个体的全部测量仍保持时间数据中的相关性; 第二, 这个方法不要求构建个体内部的相关结构.

4. 大样本性质

下面我们建立 $\hat{\beta}_{\rm LSB}(t)$ 的相合性和收敛速度. 为此, 首先引入一个距离度量来评估上述最小二乘基估计的特性. 设 $||a||_{L_2}=\left\{\int_{\mathcal{T}}[a(t)]^2{\rm d}t\right\}^{1/2}$ 是 $\mathcal{T}$ 上任一平方可积实值函数的 L_2 模, $||A||_{L_2}=\left\{\sum_{l=0}^{p}||a||_{L_2}^2\right\}^{1/2}$ 是 $A(t)=(a_0(t),\cdots,a_p(t))^{\rm T}$ 的 L_2 模, 其中 $a_l(t)$ 是 $\mathcal{T}$ 上的实值函数. 定义 $\hat{\beta}_l(t)$ 的积分平方误 (integrated squared error) 为

$$
{\rm ISE}(\hat{\beta}_l)=||\hat{\beta}_l-\beta_l||_{L_2}^2=\int_{\mathcal{T}}[\hat{\beta}_l(t)-\beta_l(t)]^2{\rm d}t,
$$

且 $\hat{\beta}_{\rm LSB}(t)=(\hat{\beta}_0(t),\cdots,\hat{\beta}_p(t))^{\rm T}$ 的积分平方误为

$$
{\rm ISE}(\hat{\beta}_{\rm LSB})=\sum_{l=0}^{p}{\rm ISE}(\hat{\beta}_l).
$$

如果 ${\rm ISE}(\hat{\beta}_{\rm LSB})\xrightarrow{P}0$, 或等价地 ${\rm ISE}(\hat{\beta}_l)\xrightarrow{P}0$, $l=0,\cdots,p$, 则称 $\hat{\beta}_{\rm LSB}(t)$ 是 $\beta(t)$ 的相合估计.

因为我们在一个线性空间中使用函数来逼近 $\beta_l(t)$, 所以 ${\rm ISE}(\hat{\beta}_{\rm LSB})$ 的渐近性依赖于 $\beta_l(t)$ 与所选的线性空间之间的某个 L_∞ 距离. 特别地, 设 $\mathcal{C}_l$ 是由 $\{B_{l1}(t),\cdots,B_{lK_l}\}$ 生成的线性空间, 并设 $D(\beta_l,\mathcal{C}_l)=\inf\limits_{g\in\mathcal{C}_l}\sup\limits_{t\in\mathcal{T}}|\beta_l(t)-g(t)|$ 是 $\beta_l(t)$ 和 $\mathcal{C}_l$ 间的 L_∞ 距离. 那么, ${\rm ISE}(\hat{\beta}_{\rm LSB})$ 的渐近性质依赖于 $\rho_n=\sum_{l=0}^{p}D(\beta_l,\mathcal{C}_l)$,

$$
A_{n,l}=\sup_{g\in\mathcal{C}_l,||g||_{L_2}\neq 0}\frac{\sup\limits_{t\in\mathcal{T}}|g(t)|}{||g||_{L_2}},\quad A_n=\max_{0\leqslant l\leqslant p}A_{n,l}.
$$

对于通常使用的诸如多项式、样条和三角矩阵基, ρ_n 和 A_n 的例子可以在 (Huang, 1998, §2.2) 中找到.

假设观测时间点是随机设计且在一个有限区间 $\mathcal{T}$ 上取值, 即 $\{t_{ij},1\leqslant i\leqslant n,1\leqslant j\leqslant m_i\}$ 独立地取自未知分布 $F(t)$, 且 $F(t)$ 具有密度 $f(t)$. 记 $K_n=\max\limits_{0\leqslant l\leqslant p}K_l$. 当 $n\to\infty$, K_n 可趋于无穷, 也可不趋于无穷. 下面两个定理给出了 $\hat{\beta}_{\rm LSB}(t)$ 的相合性和渐近正态性, 其证明可参见 (Huang et al., 2002).

定理5.3.2 设① 存在正的常数 C_1 和 C_2, 使得对任何 $t\in\mathcal{T}$ 都有 $C_1\leqslant f(t)\leqslant C_2$; ② 存在正的常数 C_3 和 C_4, 使得对任何 $t\in\mathcal{T}$ 都有 $C_3\leqslant\lambda_l(t)\leqslant C_4$, $l=0,\cdots,p$, 其中 $\lambda_0(t)\leqslant\cdots\leqslant\lambda_p(t)$ 是 $E[X(t)X^{\rm T}(t)]$ 的特征根; ③ 存在正的常数

C_5 使得对任何 $t \in \mathcal{T}$ 都有 $|X_l(t)| \leqslant C_5$, $l = 0, \cdots, p$; ④ 存在正的常数 C_6, 使得对任何 $t \in \mathcal{T}$ 都有 $E[\varepsilon^2(t)] \leqslant C_6$; ⑤ $\lim_{n\to\infty} \rho_n = 0$, 且

$$\lim_{n\to\infty} \left(A_n^2 K_n \max\left\{ \max_{1\leqslant i\leqslant n} (m_i w_i),\ \sum_{i=1}^n m_i^2 w_i^2 \right\} \right) = 0.$$

则 $\hat{\beta}_{\mathrm{LSB}}(t)$ 以概率 1 唯一存在且是 $\beta(t)$ 的相合估计, 而且

$$\mathrm{ISE}(\hat{\beta}_{\mathrm{LSB}}) = O_P\left(K_n \sum_{i=1}^n m_i^2 w_i^2 + \rho_n^2 \right).$$

注意到定理 5.3.2 对一般的基选择给出了 $\hat{\beta}_{\mathrm{LSB}}(t)$ 的相合性, 其中包括多项式、样条和三角矩阵基. 然而, 当使用特殊类型的基时, 可以改进它的收敛速度. 对一个非常有兴趣的特别情况, 下列定理 5.3.3 改进了一类样条估计的收敛速度.

定理5.3.3 假设 $\mathcal{C}_l$ 是 $\mathcal{T}$ 上具有固定度数的多项式样条空间, 连续节点之间的不同的比是有界的且大于 0. 如果定理 5.3.2 的条件成立, 则

$$\mathrm{ISE}(\hat{\beta}_{\mathrm{LSB}}) = O_P\left(\sum_{i=1}^n m_i^2 w_i^2 [(K_n/m_i) + 1] + \rho_n^2 \right).$$

注5.3.5 w_i 的不同选择一般导致估计量的不同收敛速度. 对于定理 5.3.2 的具体情况, 有

$$\sum_{i=1}^n K_n m_i^2 w_i^2 = \begin{cases} K_n/n, & w_i = 1/(nm_i), \\ K_n \sum_{i=1}^n m_i^2/N^2, & w_i = 1/N. \end{cases}$$

正如 Hoover 等 (1998) 证明的结论: $\lim_{n\to\infty} \sum_{i=1}^n m_i^2/N^2 = 0$ 当且仅当 $\lim_{n\to\infty} \max_{1\leqslant i\leqslant n} (m_i/N) = 0$. 因此, 由于利用局部光滑方法, $w_i = 1/N$ 权可能导致一个不相合的估计 $\hat{\beta}(t)$, 而 $w_i = 1/(nm_i)$ 对所有 m_i 的选择将导致相合的 $\hat{\beta}(t)$.

注5.3.6 当给定特殊的光滑条件时, 可以通过确定 $D(\beta_l, \mathcal{C}_l)$ 的大小而得到更精确的收敛速度. 例如, 当 $\beta_l(t)$ 有有界的二阶导数, $\mathcal{C}_l$ 是 $\mathcal{T}$ 上具有 K_n 个内节点的立方样条空间, 此时 $D(\beta_l, \mathcal{C}_l) = O(K_n^{-2})$(Schumaker, 1981, 定理 6.27), 由定理 5.3.2 得到 $\mathrm{ISE}(\hat{\beta}_{\mathrm{LSB}}) = O_P(K_n/n + K_n^{-4})$. 对 $K_n = O(n^{1/5})$ 的特殊选择, 它简化为 $\mathrm{ISE}(\hat{\beta}_{\mathrm{LSB}}) = O_P(n^{-4/5})$, 这正是在相同光滑条件下具有独立同分布的非参数回归估计的最优收敛速度.

5. **假设检验**

通常的兴趣是检验一个或几个系数函数是否随时间变化或同时等于 0. 下面提出一种拟合优度检验方法. 考虑检验零假设

$$H_0 : \beta_l(t) = \beta_l^0, \ \ \forall t \in \mathcal{T}, \ 1 \leqslant l \leqslant p,$$

其中 β_l^0 是未知常数, 对立假设是一个或更多的数函数 $\beta_l(l=1,\cdots,p)$ 随时间变化. 如果我们用 $\sum\limits_{s=1}^{K_0}\gamma_{0s}B_{0s}(t)$ 逼近基准函数 $\beta_0(t)$, 那么在 H_0 下加权残差平方和为

$$\mathrm{RSS}_0 = \sum_{i=1}^{n}\sum_{j=1}^{m_i} w_i \left\{ Y_{ij} - \sum_{s=1}^{K_0}\hat{\gamma}_{0s}^0 B_{0s}(t_{ij})X_{i0}(t_{ij}) - \sum_{l=1}^{p}\sum_{s=1}^{K_l}\hat{\beta}_l^0 X_{il}(t_{ij}) \right\}^2,$$

其中 $\hat{\gamma}_{0s}^0$ 和 $\hat{\beta}_l^0$ 为这个加权残差平方和的最小值点. 在对立假设下允许所有的系数函数随时间变化, 因此相应的加权残差平方和为

$$\mathrm{RSS}_1 = \sum_{i=1}^{n}\sum_{j=1}^{m_i} w_i \left\{ Y_{ij} - \sum_{l=0}^{p}\sum_{s=1}^{K_l}\hat{\gamma}_{ls} B_{ls}(t_{ij})X_{il}(t_{ij}) \right\}^2.$$

构造检验统计量

$$T_n = \frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}.$$

我们有下列结果

定理5.3.4 假设定理 5.3.2 的条件成立, $\inf\limits_{t\in\mathcal{T}}\sigma^2(t) > 0$, $\sup\limits_{t\in\mathcal{T}} E[\varepsilon^4(t)] < \infty$, 则在 H_0 下, $T_n \xrightarrow{P} 0$. 否则, 如果对某个 $1 \leqslant l \leqslant p$, $\inf\limits_{c\in R}||\beta_l - c|| > 0$, 则存在 $\delta > 0$, 使得 $P(T_n > \delta) \longrightarrow 1$.

这个结果给出了检验 H_0 的方法: 当 T_n 大于一个适当的准则值时拒绝 H_0. 我们使用 bootstrap 再抽样方法估计 T_n 的零分布. 为此, 记

$$\hat{\varepsilon}_{ij} = Y_{ij} - \sum_{l=0}^{p}\sum_{s=1}^{K_l}\hat{\gamma}_{ls}B_{ls}(t_{ij})X_{il}(t_{ij}),$$

$$\breve{Y}_{ij} = \sum_{s=1}^{K_0}\hat{\gamma}_{0s}^0 B_{0s}(t_{ij})X_{i0}(t_{ij}) - \sum_{l=1}^{p}\sum_{s=1}^{K_l}\hat{\beta}_l^0 X_{il}(t_{ij}) + \hat{\varepsilon}_{ij}.$$

下列 bootstrap 程序可以用来估计 T_n 的零分布和检验的 p 值.

步骤 1. 抽取来自 $\{(t_{ij}, X_i(t_{ij}), \breve{Y}_{ij}), 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m_i\}$ 的替代的 n 个个体来得到 bootstrap 样本 $\{(t_{ij}^*, X_i^*(t_{ij}^*), \breve{Y}_{ij}^*), 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m_i^*\}$.

步骤 2. 重复上面的再抽样过程 B 次.

步骤 3. 基于 B 个 bootstrap 样本, 用每个 bootstrap 样本计算检验统计量 T_n^*, 并得到 T_n^* 的经验分布.

步骤 4. 当 $T_n > c_\alpha^*$ 时, 则以显著性水平 α 拒绝 H_0, 其中 c_α^* 为 T_n^* 的经验分布的 $1-\alpha$ 分位点. 检验的 p 值是 $\{T_n^* > T_n\}$ 的经验概率.

为计算方便, 在 T_n^* 和 T_n 的计算中每一个 β_l 用同样的基系. 然而, 原则上不同的基逼近可以用不同的 bootstrap 复制, 这样做将增加计算量.

可以简单修改上述检验程序来检验诸如系数函数的一个或一个子集为常数或同时为 0. 定理 5.3.2 可以适应一般情况. 提出的检验方法也可以用于其他估计方法, 如局部多项式拟合和光滑样条.

5.4　经验似然推断

本节主要用经验似然方法研究模型 (5.3.1), 提出了对经验似然比纠偏的思想和系数函数置信带的构造方法. 本节提出的方法对截面数据分析仍使用. 本节的主要内容来自 (Xue and Zhu, 2007).

5.4.1　自然的经验似然

为了清楚地获得 $\beta(t)$ 的置信域构造的论据, 我们从最小二乘法的描述开始. 我们可以极小化给定 $t \in \mathbf{R}$ 下的条件均方误差 $E\{[Y(t)-\beta^{\mathrm{T}}(t)X(t)]^2|t\}$ 的样本版本定义 $\beta(t)$ 最小二乘估计, 或者解方程 $E\{[Y(t)-\beta^{\mathrm{T}}(t)X(t)]X(t)|t\}=0$. 因为与给定 t 的条件期望有关, 需要局部光滑方法得到样本版本. 为定义经验似然比函数, 可利用约束 $E\{[Y(t)-\beta^{\mathrm{T}}(t)X(t)]X(t)|t\}=0$. 为此, 引入辅助随机向量

$$Z_i(\beta(t)) = \sum_{j=1}^{m_i}\left[Y_{ij}-\beta^{\mathrm{T}}(t)X_i(t_{ij})\right]X_i(t_{ij})K\left(\frac{t_{ij}-t}{h}\right), \tag{5.4.1}$$

其中 h 是带宽, $K(\cdot)$ 是核函数.

注意到 $\{Z_i(\beta(t)), 1\leqslant i\leqslant n\}$ 是独立的, 且 $E[Z_i(\beta(t))]=0$. 因此, $\beta(t)$ 的一个自然的经验对数似然比函数可定义为

$$\mathcal{R}(\beta(t)) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\Big|\, p_i\geqslant 0, \sum_{i=1}^{n}p_i=1, \sum_{i=1}^{n}p_iZ_i(\beta(t))=0\right\},$$

其中 $p_i=p_i(t), i=1,\cdots,n$. 对于给定的函数系数 $\beta(t)$, $\mathcal{R}(\beta(t))$ 的单位值存在, 假若 0 在点 $(Z_1(\beta(t)),\cdots,Z_n(\beta(t)))$ 的凸零集内部 (Owen, 1988, 1990). 由 Lagrange

乘子法, 可以把 $\mathcal{R}(\beta(t))$ 表示为

$$\mathcal{R}(\beta(t))=2\sum_{i=1}^{n}\log\left(1+\theta^{\mathrm{T}}Z_i(\beta(t))\right), \tag{5.4.2}$$

其中 θ 是 $(p+1)\times 1$ 向量且满足

$$\sum_{i=1}^{n}\frac{Z_i(\beta(t))}{1+\theta^{\mathrm{T}}Z_i(\beta(t))}=0. \tag{5.4.3}$$

使用 (5.4.3) 式和 Taylor 展式可以证明: 对一个固定点 $t_0\in\mathbf{R}$, 如果 $\beta(t_0)$ 是真参数, 则

$$\mathcal{R}(\beta(t_0))=\left[\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}Z_i(\beta(t_0))\right]^{\mathrm{T}}\widetilde{D}^{-1}(\beta(t_0))\left[\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}Z_i(\beta(t_0))\right]+o_P(1), \tag{5.4.4}$$

其中

$$\widetilde{D}(\beta(t_0))=\frac{1}{Nh}\sum_{i=1}^{n}Z_i(\beta(t_0))Z_i^{\mathrm{T}}(\beta(t_0)). \tag{5.4.5}$$

因此, 在适当的条件下 $\mathcal{R}(\beta(t_0))$ 是渐近 χ^2 的. $\mathcal{R}(\beta(t_0))$ 的渐近性质如下.

定理5.4.1 设条件 (C5.3.2)~(C5.3.7) 成立, 且 $Nh\to\infty$, $Nh^5\to 0$. 如果 $\beta(t_0)$ 是真参数, 则 $\mathcal{R}(\beta(t_0))\xrightarrow{D}\chi^2_{p+1}$.

可以求得 $\mathcal{R}(\beta(t))$ 极小值点, 记为 $\hat{\beta}(t)$, 称它为 $\beta(t)$ 的极大经验似然估计 (MELE). 由 (5.4.2) 和 (5.4.3) 式可以证明: 估计量 $\hat{\beta}(t)$ 是估计方程 $\sum\limits_{i=1}^{n}Z_i(\beta(t))=0$ 的解. 通过解该估计方程可以得到 $\hat{\beta}(t)=\hat{\beta}_{\mathrm{LK}}(t)$, 其中 $\hat{\beta}_{\mathrm{LK}}(t)$ 是由 Wu 等 (1998) 得到的局部核估计 (LKE), 见 (5.3.5) 式. 这说明 MELE 与 LKE 是等价的, 即, MELE 与 LKE 有相同的渐近分布. 因此, MELE 也具有定理 5.3.1 给出的渐近正态性.

5.4.2 两种纠偏方法

在定理 5.4.1 中, 带宽 h 的范围在区间 $(c_1N^{-1/2},c_2N^{-1/5})$ 内部, 其中 $c_1>0$ 和 $c_2>0$ 是两个常数. 因为 h 的取值不包括最优带宽, 因此需要用欠光滑消除偏差. 然而, 这就涉及如何恰当地选择带宽, 常常会有一定难度. 为了解决这个问题, 我们提出两种改进的经验似然比.

1. 均值校正的经验似然

由 Wu 等 (1998) 的引理 A1 可以证明：当条件 (C5.3.1)~(C5.3.7) 成立时,

$(Nh)^{-1/2}\sum_{i=1}^{n} Z_i(\beta(t_0))$ 渐近服从非零均值向量 $b(t_0)$ 和协方差阵 $\upsilon^2(t_0)\Gamma(t_0)$ 的正态分布. 由 (5.4.4) 可知 $\mathcal{R}(\beta(t_0))$ 不是渐近标准 χ^2 变量. 因此, 一个校正的自然方法是借助扣除与 $\mathrm{b}(t_0)$ 有关的估计量. 这个想法有点像 Wu 等 (1998) 曾使用的正态逼近和中心化统计量的方法. 该统计思想的详细叙述如下. 我们首先估计 $b_l(t_0)$, $l=0,\cdots,p$. 可以证明

$$E\{X_{il}(t_{ij})X_i^{\mathrm{T}}(t_{ij})[\beta(t_{ij})-\beta(t_0)]K((t_0-t_{ij})/h)\}(N/h)^{1/2}=b_l(t_0)+o(1).$$

因此, $b_l(t_0)$ 的一个相合估计定义为

$$\hat{b}_l(t_0)=\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}X_{il}(t_{ij})X_i^{\mathrm{T}}(t_{ij})[\hat{\beta}(t_{ij})-\hat{\beta}(t_0)]K\left(\frac{t_{ij}-t_0}{h}\right). \tag{5.4.6}$$

$b(t_0)$ 的估计量定义为 $\hat{b}(t_0)=(\hat{b}_0(t_0),\cdots,\hat{b}_p(t_0))$. 记 $\tilde{\zeta}(\beta(t_0))=(Nh)^{-1/2}\sum_{i=1}^{n}Z_i(\beta(t_0))$, $\tilde{\xi}(\beta(t_0))=\hat{b}^{\mathrm{T}}(t_0)\widetilde{D}^{-1}(\beta(t_0))[2\tilde{\zeta}(\beta(t_0))-\hat{b}(t_0)]$, 其中 $\widetilde{D}(\beta(t_0))$ 在 (5.4.5) 式中定义. 因此, $\beta(t_0)$ 的一个均值校正的经验似然比函数定义为

$$\widetilde{\mathcal{R}}(\beta(t_0))=\mathcal{R}(\beta(t_0))-\tilde{\xi}(\beta(t_0)). \tag{5.4.7}$$

下列结果说明了 $\widetilde{\mathcal{R}}(\beta(t_0))$ 是渐近 χ^2 的.

定理5.4.2　设条件 (C5.3.1)~(C5.3.7) 成立, 且核 $K(\cdot)$ 在 t_0 点二次连续可微. 如果 $\beta(t_0)$ 是真参数, 则 $\widetilde{\mathcal{R}}(\beta(t_0))\xrightarrow{D}\chi^2_{p+1}$.

2. *残差调整的经验似然*

对经验似然比的一个奏效的修正是利用它自身的渐近表达来实现. 通过对自然经验似然比渐近性质的推证发现, 对加权残差 $Z_i(\beta(t))$ 进行调整将有助于减少偏差. 为此, 引入辅助随机向量

$$\hat{Z}_i(\beta(t))=\sum_{j=1}^{m_i}\{Y_{ij}-\beta^{\mathrm{T}}(t)X_i(t_{ij})-[\hat{\beta}(t_{ij})-\hat{\beta}(t)]^{\mathrm{T}}X_i(t_{ij})\}X_i(t_{ij})K_h(t_{ij}-t).$$

显然, $\hat{Z}_i(\beta(t))$ 是 (5.4.1) 式中 $Z_i(\beta(t))$ 的调整. $\beta(t)$ 的残差调整的经验对数似然比定义为

$$\hat{\mathcal{R}}(\beta(t))=-2\max\left\{\sum_{i=1}^{n}\log(np_i)\middle|p_i\geqslant 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\hat{Z}_i(\beta(t))=0\right\}.$$

$\hat{\mathcal{R}}(\beta(t))$ 的结果陈述在下列定理中.

定理5.4.3 设条件 (C5.3.1)~(C5.3.7) 成立, 且核 $K(t)$ 在点 t_0 处是二次可微的. 如果 $\beta(t_0)$ 是真参数, 则 $\hat{\mathcal{R}}(\beta(t_0)) \xrightarrow{D} \chi^2_{p+1}$.

5.4.3 近似置信域

本节考虑三个问题: 基于经验似然, $\beta(t_0)(\beta_r(t_0))$ 的逐点置信域 (区间) 的构造; 特别感兴趣分量的同时置信带的构造; 基于 MELE(LKE) 正态逼近的逐点和同时置信域的构造.

1. 一般情况

使用定理 5.4.1~定理 5.4.3, 并基于自然、均值校正和残差调整的经验似然比, 可以构造 $\beta(t_0)$ 的近似 $1-\alpha$ 置信域即

$$R_\alpha(t_0) = \left\{\tilde{\beta}(t_0) | \mathcal{R}(\tilde{\beta}(t_0)) \leqslant \chi^2_{p+1}(1-\alpha)\right\},$$
$$\widetilde{R}_\alpha(t_0) = \left\{\tilde{\beta}(t_0) | \widetilde{\mathcal{R}}(\tilde{\beta}(t_0)) \leqslant \chi^2_{p+1}(1-\alpha)\right\},$$
$$\hat{R}_\alpha(t_0) = \left\{\tilde{\beta}(t_0) | \hat{\mathcal{R}}(\tilde{\beta}(t_0)) \leqslant \chi^2_{p+1}(1-\alpha)\right\}.$$

2. 分量的逐点置信区间

如果对 $\beta(t)$ 的某一分量的逐点置信区间特别感兴趣, 那么可以利用部分 profile 经验似然比构造其置信区间. 为此, 用 e_{r+1} 表示长度为 $p+1$ 且第 $r+1$ 个位置为 1 的单位向量, $r=0,\cdots,p$. 由 (5.3.5) 式, 分量 $\beta_r(t)$ 的估计量是 $\hat{\beta}_r(t)=e^{\mathrm{T}}_{r+1}\hat{\beta}_{\mathrm{LK}}(t)$, $r=0,\cdots,p$. 记

$$Z_{i,r}(\beta_r(t)) = e^{\mathrm{T}}_{r+1}\hat{V}^{-1}(t)Z_i(\hat{\beta}_0(t),\cdots,\hat{\beta}_{r-1}(t),\beta_r(t),\hat{\beta}_{r+1}(t),\cdots,\hat{\beta}_p(t)),$$
$$\hat{Z}_{i,r}(\beta_r(t)) = e^{\mathrm{T}}_{r+1}\hat{V}^{-1}(t)\hat{Z}_i(\hat{\beta}_0(t),\cdots,\hat{\beta}_{r-1}(t),\beta_r(t),\hat{\beta}_{r+1}(t),\cdots,\hat{\beta}_p(t)).$$

类似 (5.4.2) 式, $\beta_r(t)$ 的自然的和残差调整的经验对数似然比函数分别定义为

$$\mathcal{R}_r(\beta_r(t)) = 2\sum_{i=1}^n \log\left(1+\theta_1 Z_{i,r}(\beta_r(t))\right),$$

$$\hat{\mathcal{R}}_r(\beta_r(t)) = 2\sum_{i=1}^n \log\left(1+\theta_2 \hat{Z}_{i,r}(\beta_r(t))\right),$$

其中 θ_1 和 θ_2 如同 (5.4.3) 式给定.

在定理 5.4.1 和定理 5.4.2 的假定下, 我们可以证明 $\mathcal{R}_r(\beta_r(t_0))$ 和 $\hat{\mathcal{R}}_r(\beta_r(t_0))$ 的渐近分布是自由度为 1 的 χ^2 分布. 这是一个有趣的结果, 因为在这种情况下, 其他分量的估计并不减弱 χ^2 收敛性. 下面的定理陈述了这个结果.

定理5.4.4 在定理 5.4.1 的条件下, 有 $\mathcal{R}_r(\beta_r(t_0)) \xrightarrow{D} \chi^2_1$. 当定理 5.4.2 的条件满足时, 有 $\hat{\mathcal{R}}_r(\beta_r(t_0)) \xrightarrow{D} \chi^2_1$.

应用定理 5.4.4, 可以定义 $\beta_r(t_0)$ 的近似 $1-\alpha$ 置信区间为

$$
\begin{aligned}
I_{r,\alpha}(t_0) &= \{\tilde{\beta}_r(t_0) |\, \mathcal{R}_r(\tilde{\beta}_r(t_0)) \leqslant \chi_1^2(1-\alpha)\},\\
\hat{I}_{r,\alpha}(t_0) &= \{\tilde{\beta}_r(t_0) |\, \hat{\mathcal{R}}_r(\tilde{\beta}_r(t_0)) \leqslant \chi_1^2(1-\alpha)\}.
\end{aligned} \tag{5.4.8}
$$

我们也可以通过改良由 (5.4.7) 式定义的均值校正的经验对数似然比来构造任何分量 $\beta_r(t_0)$ 的置信区间. 记

$$
\widetilde{Z}_r(\beta_r(t_0)) = (Nh)^{-1/2}\sum_{i=1}^{n} Z_{i,r}(\beta_r(t_0)),
$$

$$
\widetilde{D}_r(\beta_r(t_0)) = (Nh)^{-1}\sum_{i=1}^{n} Z_{i,r}^2(\beta_r(t_0)),
$$

$$
\tilde{\xi}_r(\beta_r(t_0)) = \hat{b}_r(t_0)\widetilde{D}_r^{-1}(\beta_r(t_0))\left[2\widetilde{Z}_r(\beta_r(t_0)) - \hat{B}_r(t_0)\right].
$$

那么, $\beta_r(t_0)$ 的均值校正的经验对数似然比可以定义为

$$
\widetilde{\mathcal{R}}_r(\beta_r(t_0)) = \mathcal{R}_r(\beta_r(t_0)) - \tilde{\xi}_r(\beta_r(t_0)),
$$

且有下列结果.

定理5.4.5　在定理 5.4.1 的条件下, 有 $\widetilde{\mathcal{R}}_r(\beta_r(t_0)) \xrightarrow{D} \chi_1^2$.

因此, $\beta_r(t_0)$ 的一个近似 $1-\alpha$ 置信区间可定义为

$$
\tilde{I}_{r,\alpha}(t_0) = \{\tilde{\beta}_r(t_0) |\, \widetilde{\mathcal{R}}_r(\tilde{\beta}_r(t_0)) \leqslant \chi_1^2(1-\alpha)\}.
$$

注5.4.1　上述方法可以用来构造 $\beta(t_0)$ 的任何两个不同分量 $(\beta_r(t_0), \beta_s(t_0))$ 的置信域. 基本思想如下: 在 $Z_i(\beta(t_0))$ 中除 $\beta_r(t_0)$ 和 $\beta_s(t_0)$ 外, 首先用其估计量代替 $\beta(t_0)$ 的其他分量而得到 $Z_{i,(r,s)}(\beta_r(t_0), \beta_s(t_0))$; 然后得到使用

$$
(e_{r+1}, e_{s+1})^{\mathrm{T}}\hat{V}^{-1}(t_0)Z_{i,(r,s)}(\beta_r(t_0), \beta_s(t_0))
$$

代替 (5.4.2) 式中的 $Z_i(\beta(t_0))$ 而得到经验对数似然比函数 $\mathcal{R}_{r,s}(\beta_r(t_0), \beta_s(t_0))$; 最后, 使用 $\mathcal{R}_{r,s}(\beta_r(t_0), \beta_s(t_0))$ 可以构造 $(\beta_r(t_0), \beta_s(t_0))$ 的经验似然置信域. 用类似地方式可以定义 $(\beta_r(t_0), \beta_s(t_0))$ 的均值校正和残差调整的经验似然置信域. 这个方法可以容易地推广到 $\beta(t_0)$ 的若干个分量的置信域的构造.

3. 同时置信带

下面介绍将上述逐点置信区间扩充为 $\beta_r(t)$ 的同时置信带的简单方法, 这里在 $S(f)$ 的一个给定的子区间 $[a,b]$ 上考虑这个问题. 基本思想是在 $[a,b]$ 的每一个选

出的格子点上来调用逐点置信区间. 可以用 Bonferroni 方法调整这些置信区间的置信度而得到一致置信带. 可以利用回归曲线的光滑条件桥接格子点之间的间隙. 参见 (Knafl et al., 1985; Hall and Titterington, 1988; Härdle, 1989; Wu et al., 1998; Xue and Zhu, 2007) 等.

我们首先基于残差调整的经验似然考虑置信带. 将 (5.4.8) 式定义的置信区间记为 $\hat{I}_{r,\alpha}(t_0) = [\hat{L}_{r,\alpha}(t_0), \hat{U}_{r,\alpha}(t_0)]$. 由定理 5.4.4 可得

$$\lim_{n\to\infty} P\{\hat{L}_{r,\alpha}(t_0) \leqslant \beta_r(t_0) \leqslant \hat{U}_{r,\alpha}(t_0)\} = 1-\alpha.$$

设 M 是一个正数, $\gamma = (b-a)/M$, 又设 $a = \xi_1 < \cdots < \xi_{M+1} = b$ 为 $M+1$ 个格子点, 使得 $\xi_{j+1} - \xi_j = \gamma$, $j = 1, \cdots, M$. $\beta_r(\xi_j)$ 的一个近似同时置信区间记为 $[\hat{l}_{r,\alpha}(\xi_j), \hat{u}_{r,\alpha}(\xi_j)]$, 使得

$$\lim_{n\to\infty} P\{\hat{l}_{r,\alpha}(\xi_j) \leqslant \beta_r(\xi_j) \leqslant \hat{u}_{r,\alpha}(\xi_j), \forall\, j = 1, \cdots, M+1\} \geqslant 1-\alpha.$$

一个基于 Bonferroni 调整的方法是选择包含 $[\hat{L}_{r,\alpha/(M+1)}(\xi_j), \hat{U}_{r,\alpha/(M+1)}(\xi_j)]$ 的 $[\hat{l}_{r,\alpha}(\xi_j), \hat{u}_{r,\alpha}(\xi_j)]$.

对 $t \in [\xi_j, \xi_{j+1}]$, 设 $\beta_r^{(I)}(t)$ 是 $\beta_r(\xi_j)$ 和 $\beta_r(\xi_{j+1})$ 的线性内插, 即

$$\beta_r^{(I)}(t) = (\xi_{j+1} - t)\gamma^{-1}\beta_r(\xi_j) + (t - \xi_j)\gamma^{-1}\beta_r(\xi_{j+1}).$$

类似地, 设 $\hat{l}_{r,\alpha}^{(I)}(t)$ 和 $\hat{u}_{r,\alpha}^{(I)}(t)$ 分别是 $\hat{l}_{r,\alpha}$ 和 $\hat{u}_{r,\alpha}$ 在 ξ_j 和 ξ_{j+1} 处的值的线性内插. 由此得出

$$\lim_{n\to\infty} P\{\hat{l}_{r,\alpha}^{(I)}(t) \leqslant \beta_r^{(I)}(t) \leqslant \hat{u}_{r,\alpha}^{(I)}(t), \forall\, t \in [a,b]\} \geqslant 1-\alpha.$$

为桥接 $\beta_r(t)$ 和 $\beta_r^{(I)}(t)$ 之间的间隙, 必须指定 $\beta_r(t)$ 的光滑条件. 假定由 Knafl 等 (1985) 提出的下列两个条件:

$$\sup_{t\in[a,b]} |\beta_r'(t)| \leqslant c_1, \tag{5.4.9}$$

$$\sup_{t\in[a,b]} |\beta_r''(t)| \leqslant c_2, \tag{5.4.10}$$

其中 $c_1 > 0$ 和 $c_2 > 0$ 是给定的常数. 如果 (5.4.9) 式满足, 则由 Taylor 公式可以推出, 对所有 $t \in [\xi_j, \xi_{j+1}]$,

$$|\beta_r(t) - \beta_r^{(I)}(t)| \leqslant c_1\gamma^{-1}(\xi_{j+1} - t)(t - \xi_j).$$

类似地, 如果 (5.4.10) 式满足, 则可以证明, 对所有 $t \in [\xi_j, \xi_{j+1}]$,

$$|\beta_r(t)-\beta_r^{(I)}(t)|\leqslant \frac{1}{2}c_2(\xi_{j+1}-t)(t-\xi_j).$$

如果调整 $\beta_r^{(I)}(t)$ 的带, 那么在 $t\in[a,b]$ 上 $\beta_r(t)$ 的近似 $1-\alpha$ 置信带是

$$[\hat{l}_{r,\alpha}^{(I)}(t)-c_1\gamma^{-1}(\xi_{j+1}-t)(t-\xi_j),\hat{u}_{r,\alpha}^{(I)}(t)+c_1\gamma^{-1}(\xi_{j+1}-t)(t-\xi_j)] \qquad (5.4.11)$$

或

$$[\hat{l}_{r,\alpha}^{(I)}(t)-\frac{1}{2}c_2(\xi_{j+1}-t)(t-\xi_j),\hat{u}_{r,\alpha}^{(I)}(t)+\frac{1}{2}c_2(\xi_{j+1}-t)(t-\xi_j)], \qquad (5.4.12)$$

$t\in[\xi_j,\xi_{j+1}]$, $j=1,\cdots,M$.

类似地, 对自然和均值校正的经验似然比, 我们也可以构造同时置信带. 这里省略其详细内容.

注5.4.2 一个重要的问题是由 (5.4.11) 和 (5.4.12) 式给出的置信带中 M 的选择. 对独立截面数据的核回归的简单情况, 虽然 Hall 和 Titterington(1988) 提供了选择 M 的启发性建议, 即 $M>(b-a)/M$, 但在现行情况下 M 的选择仍是公开问题, 因此 M 的最优选择也需要进一步研究.

4. 基于正态逼近的置信域

逐点置信区间和置信带. 应用定理 5.3.1, 可以构造 $\beta(t_0)$ 的逐点置信区间. 但我们需要使用 $\hat{\beta}(t_0)$ 的渐近偏差和方差的嵌入估计. 显然, $\hat{\beta}(t_0)$ 的渐近偏差和方差依赖于 $b(t_0)$, $f(t_0)$, $\Gamma(t_0)$ 和 $v^2(t_0)$, 其中 $\Gamma(t_0)=\big(\gamma_{lr}(t_0)\big)$ 在条件 (C5.3.7) 中定义. $b(t_0)$ 的估计量 $\hat{b}(t_0)$ 在 (5.4.6) 式中定义. 使用核光滑方法可以估计密度 $f(t_0)$ 和协方差 $\gamma_{lr}(t_0)$. 取核函数 $K(\cdot)$ 和带宽 h, 其中 h 可不同于 $\hat{\beta}(t_0)$ 的带宽. 假设 $K(\cdot)$ 满足条件 (C5.3.3). 如果 h 满足 $h\to 0$, $Nh\to\infty$, 那么 $f(t_0)$ 和 $\gamma_{lr}(t_0)$ 的估计量可分别定义为

$$\hat{f}(t_0)=\frac{1}{Nh}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K\left(\frac{t_{ij}-t_0}{h}\right),$$

$$\hat{\gamma}_{lr}(t_0)=\frac{1}{Nh\hat{f}(t_0)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}X_{il}(t_{ij})X_{ir}(t_{ij})K\left(\frac{t_{ij}-t_0}{h}\right).$$

设 $\hat{\Gamma}(t_0)$ 是 $(p+1)\times(p+1)$ 矩阵, 其 (l,r) 元素为 $\hat{\gamma}_{lr}(t_0)$. 设 $\Gamma(t_0)$ 可逆. 那么, 可以用 $\hat{\Gamma}(t_0)$ 和 $\hat{\Gamma}^{-1}(t_0)$ 估计 $\Gamma(t_0)$ 和 $\Gamma^{-1}(t_0)$.

下面我们讨论 $v^2(t_0)$ 的估计. 由 (5.3.7) 式, $v^2(t_0)$ 包含方差 $\sigma^2(t_0)$ 和协方差 $\rho_\varepsilon(t_0)$. Wu 等 (1998) 定义了 $\sigma^2(t_0)$ 和 $\rho_\varepsilon(t_0)$ 的估计量 $\hat{\sigma}^2(t_0)$ 和 $\hat{\rho}_\varepsilon(t_0)$, 这里, 我们提出 $v^2(t_0)$ 的一个简单估计量:

$$\hat{v}^2(t_0)=\frac{1}{Nh}\sum_{i=1}^{n}\left\{\sum_{j=1}^{m_i}[Y_{ij}-X_i^{\mathrm{T}}(t_{ij})\hat{\beta}(t_0)]K\left(\frac{t_{ij}-t_0}{h}\right)\right\}^2. \qquad (5.4.13)$$

最后, 用 $\hat{f}(t_0)$, $\hat{\Gamma}^{-1}(t_0)$ 和 $\hat{b}(t_0)$ 代替 $B(t_0)$ 的 $f(t_0)$, $\Gamma^{-1}(t_0)$ 和 $b(t_0)$ 而得到估计量 $\hat{B}(t_0)$; 用 $\hat{f}(t_0)$, $\hat{v}^2(t_0)$ 和 $\hat{\Gamma}^{-1}(t_0)$ 代替 $\Sigma(t_0)$ 的 $f(t_0)$, $v^2(t_0)$ 和 $\Gamma^{-1}(t_0)$ 而得到估计量 $\hat{\Sigma}(t_0)$.

下列命题给出了 $\hat{b}(t_0)$ 和 $\hat{v}^2(t_0)$ 的相合性.

命题5.4.1 设条件 (C5.3.1)~(C5.3.6) 成立, 则 $\hat{b}(t_0) \xrightarrow{P} b(t_0)$, $\hat{f}(t_0) \xrightarrow{P} f(t_0)$, $\hat{\gamma}_{lr} \xrightarrow{P} \gamma_{lr}(t_0)$, $\hat{v}^2(t_0) \xrightarrow{P} v^2(t_0)$.

由命题 5.4.1 可知 $\hat{B}(t_0)$ 和 $\hat{\Sigma}(t_0)$ 分别是 $B(t_0)$ 和 $\Sigma(t_0)$ 的相合估计. 由定理 5.3.1 可得

$$\hat{\Sigma}^{-1/2}(t_0)\left\{\sqrt{Nh}\,[\hat{\beta}(t_0) - \beta(t_0)] - \hat{B}(t_0)\right\} \xrightarrow{D} N(0, I_{p+1}), \tag{5.4.14}$$

其中 I_{p+1} 是 $p+1$ 阶单位矩阵.

使用 (5.4.14) 式, 可定义 $\beta_r(t_0)$ 的一个逐点置信区间为

$$\hat{\beta}_r(t_0) - (Nh)^{-1/2}\hat{B}_r(t_0) \pm z_{1-\alpha/2}(Nh)^{-1/2}\hat{\sigma}_r(t_0),$$

其中 $\hat{B}_r(t_0)$ 是 $\hat{B}(t_0)$ 的第 r 个分量, $\hat{\sigma}_r^2(t_0)$ 是 $\hat{\Sigma}(t_0)$ 的 (r,r) 元素, $z_{1-\alpha/2}$ 是标准正态分布的 $1-\alpha/2$ 分位数.

利用上述方法也可以构造 $\beta_r(t)$ 的近似 $1-\alpha$ 置信带. 这里省略其叙述.

置信域. 记 $\hat{S}(\beta(t_0)) = \sqrt{Nh}[\hat{\beta}(t_0) - \beta(t_0)] - \hat{B}(t_0)$. 使用 (5.4.14) 式和 (Arnold, 1981) 中的定理 10.2d 可得

$$\hat{S}^{\mathrm{T}}(\beta(t_0))\hat{\Sigma}^{-1}(t_0)\hat{S}(\beta(t_0)) \xrightarrow{D} \chi^2_{p+1}.$$

这个结果可以用来构造 $\beta(t_0)$ 的置信域以及其他分量的置信域. 例如, 对任何 $0 \leqslant r, s \leqslant k$ 和 $r \neq s$, 我们可以使用下列方法构造 $(\beta_r(t_0), \beta_s(t_0))$ 的置信域. 记

$$d(\beta_r(t_0)) = \sqrt{Nh}\,[\hat{\beta}_r(t_0) - \beta_r(t_0)] - \hat{B}_r(t_0),$$

$$U_{r,s}(t_0) = \big(d(\beta_r(t_0)), d(\beta_s(t_0))\big),$$

$$m(\beta_r(t_0), \beta_s(t_0)) = U_{r,s}^{\mathrm{T}}(t_0)(e_{r+1}, e_{s+1})^T\hat{\Sigma}^{-1}(t_0)(e_{r+1}, e_{s+1})U_{r,s}(t_0),$$

其中 e_{r+1} 是在 5.4.3 小节中定义的列向量. 那么 $(\beta_r(t_0), \beta_s(t_0))$ 的近似 $1-\alpha$ 置信域定义为

$$R_{rs,\alpha}(t_0) = \left\{(\tilde{\beta}_r(t_0), \tilde{\beta}_s(t_0)) \middle|\, m(\tilde{\beta}_r(t_0), \tilde{\beta}_s(t_0)) \leqslant \chi^2_2(1-\alpha)\right\}.$$

5.4.4 数值计算结果

1. 模拟研究

核光滑方法要求带宽选择, 主要是为了调整曲线的质量. 由于纵向数据个体之间是独立的, 因此可以利用 Rice 和 Silverman(1991) 提出的 "抛去一个体" 交叉验证方法选择带宽. 具体细节可参见 5.3.2 小节和 (5.3.6) 式, 这里不再赘述. 设 CV(h) 是使用 "抛去一个体" 交叉验证 (CV) 方法选择的带宽. 按照 Wu 等 (1998) 的带宽选择方法, 也可以利用 "抛去一个体" 交叉验证方法选择 $\hat{f}(t_0)$ 和 $\hat{\gamma}_{lr}(t_0)$ 的带宽 $h_{\mathrm{CV}(f)}$ 和 $h_{\mathrm{CV}(\gamma_{lr})}$. 然而, 正如 Wu 等 (1998) 指出, 交叉验证方法计算量很大. 因此, 为减轻计算量, 我们使用 (5.3.6) 式的 CV(h) 代替 $h_{\mathrm{CV}(f)}$ 和 $h_{\mathrm{CV}(\gamma_{lr})}$. 由 (5.4.6) 和 (5.4.13) 式定义的估计量 $\hat{b}_l(t_0)$ 和 $\hat{v}(t_0)$ 的带宽也是 h_{CV}. 基于下面我们的有限模拟将看到这样简单的带宽选择产生满意的置信域和带.

我们从具有协变量 $X_i(t) = (1, X_{i1}(t), X_{i2}(t), X_{i3}(t))$ 的模型 (5.3.1) 中产生数据, 其中 $X_{i1}(t)$ 和 $X_{i2}(t)$ 相互独立且都服从 $[t/12, 2+t/12]^2$ 上的均匀分布, $X_{i3}(t)$ 与 $X_{i1}(t)$ 和 $X_{i2}(t)$ 独立且服从正态分布 $N(0, 0.8)$. 系数函数为 $\beta_0(t) = \sin(\pi t/6)$, $\beta_1(t) = 1 - ((t-6)/6)^2$, $\beta_2(t) = \cos(\pi t/6)$, $\beta_3(t) = 0.01\exp(t-6)$. 独立地重复 200 次计算. 在每一次重复计算中, 产生 100 个个体和他们的相应协变量 $X_i(t)$, $i = 1, \cdots, 100$. 我们也产生 12 个等距的 "预定" 时间点和 100 个来自 $U(0,1)$ 分布的随机位移点 s_{i1}, 使得 $s_{il} = s_{i1} + (l-1)$, $l = 1, \cdots, 12$. 此外, 每一个预定的时间点 s_{il} 有 60% 的概率随机删失. 逗留观察的时间点记为 t_{ij}. 这就导致重复测量的不等距数目 m_i 和不同的观察时间点 t_{ij}, $i = 1, \cdots, 100$, $j = 1, \cdots, m_i$. 误差 $\varepsilon_i(t_{ij})$ 是从零均值且协方差函数为

$$\operatorname{cov}(\varepsilon_{i_1}(t_{i_1 j_1}), \varepsilon_{i_2}(t_{i_2 j_2})) = \begin{cases} 0.2\exp(-|t_{i_1 j_1} - t_{i_2 j_2}|), & i_1 = i_2, \\ 0, & i_1 \neq i_2 \end{cases}$$

的 Gauss 过程中产生. 将 $(t_{ij}, X_i(t_{ij}), \varepsilon_i(t_{ij}))$ 和前述的系数函数代入到 (5.3.1) 式中可得到依赖时间的相应变量 Y_{ij}.

对每个模拟的数据集, 我们使用标准 Gaussian 核和 h_{CV} 计算 $\beta(t)$ 的经验似然比和估计量. 从同时置信带的宽度和他们的相应覆盖概率来比较三个方法的优劣. 这三个方法是: 残差调整的经验似然 (RAEL), 均值校正的经验似然 (MCEL) 和正态逼近 (NA). 用 $M = 10$ 构造了近似 0.95 Bonferroni 类同时置信带. 模拟结果展示在图 5.4.1.

图 5.4.1 表明 RAEL 和 MCEL 比其他两个方法给出更狭窄的置信带. 对所有方法而言, 保守的 Bonferroni 带有高达接近于 1 的覆盖概率. 我们也对逐点的置信区间做了模拟研究, 对所有方法的模拟结果显示覆盖范围大约是 0.942 到 0.934,

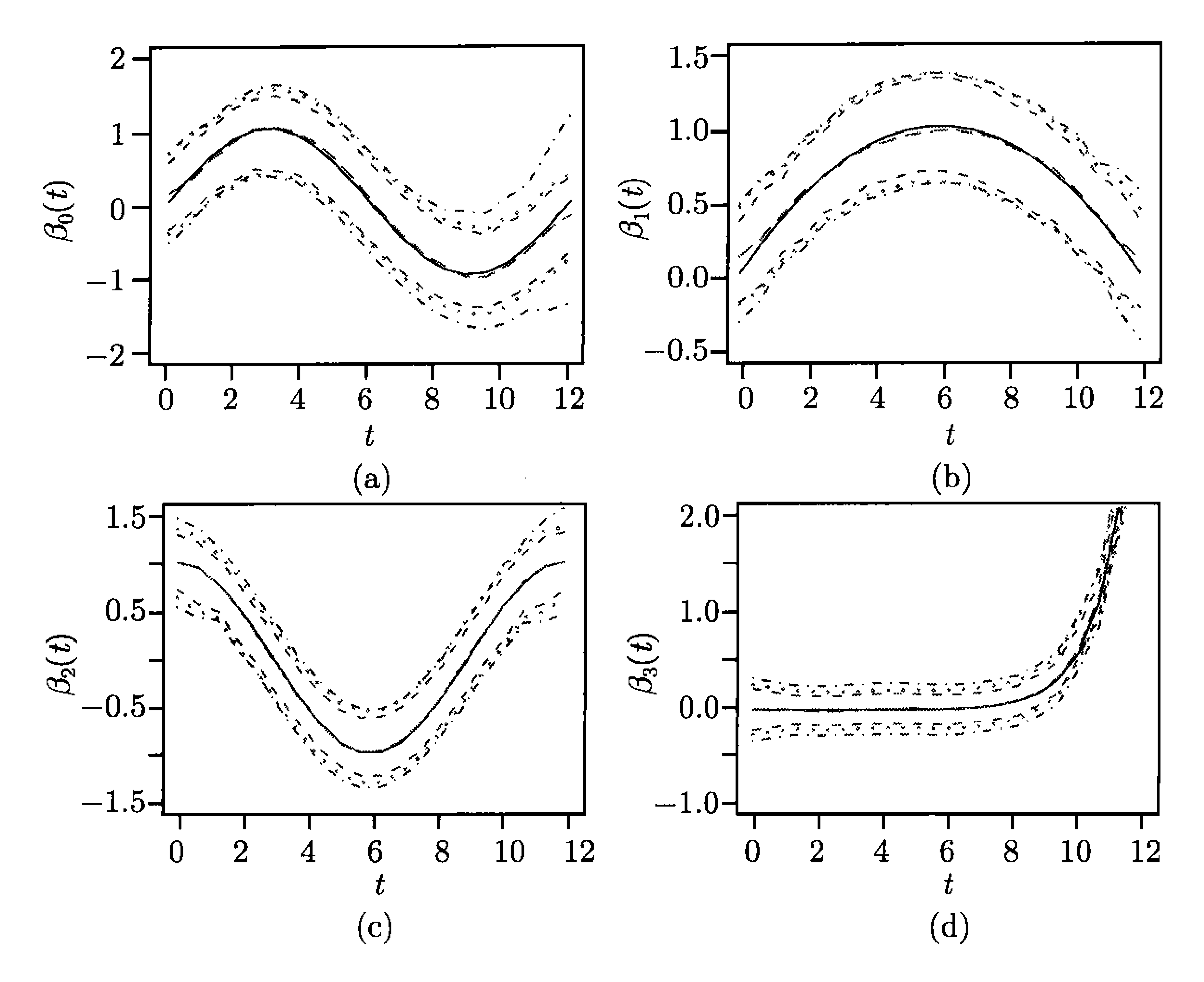


图 5.4.1 基于 RANA(虚线)、MCNA(点线) 和 NA(点虚线), $\beta_r(t)(r=0,1,2,3)$ 的近似 0.95 Bonferroni 类同时置信带. 实线是真曲线, 长虚线是 $\beta_r(t)(r=0,1,2,3)$ 的估计的曲线. 图 (a)~(d) 分别对应 $\beta_0(t)$、$\beta_1(t)$、$\beta_2(t)$ 和 $\beta_3(t)$

RAEL 和 MCEL 比其他两个方法也给出更短的置信区间, 且达到了更高的经验覆盖概率. 因此, 这说明它不是简单地对宽区间高覆盖的折中. 从模拟结果可知两个修正的经验似然比一般产生好的效果.

2. *艾滋病数据的应用*

我们现在通过分析来自艾滋病研究中心的数据来说明经验似然方法的应用. 这个数据集包含在 1984~1991 年跟踪调查期间 283 个感染 HIV 的同性恋男人的 HIV 情况. HIV 减弱人的抵抗力而导致艾滋病. HIV 损害人体内的 CD4 细胞. 未感染的人体每毫升血液中大约含 1100 个 CD4 细胞. 当人体感染 HIV 后, 其 CD4 细胞数会随着时间的推移而不断减少. 因此, 患者的 CD4 细胞数可以用来观察病情. 实验参加者每隔半年检查一次 CD4 细胞数. 然而, 由于一些人在研究期间失掉了他们的预定访问和 HIV 随机发生, 因此有不等的重复测量数和各个体的不同测量时间, 且产生的纵向数据是非平衡的.

研究的目的是描述整个时期平均 CD4 百分率消耗的趋势, 并根据感染后平均 CD4 百分率来评估在 HIV 感染下吸烟、年龄和 HIV 感染前 CD4 百分率的影响. 这个问题可以用变系数模型 (5.3.1) 来建模. 假设 t_{ij} 是感染 HIV 后第 i 个个体第

j 次测量的时间 (年度); Y_{ij} 在时间 t_{ij} 处第 i 个个体的 CD4 细胞百分率; X_{i1} 是第 i 个个体的吸烟情况, 其值取为 1 或 0 视他感染后总是或从不吸烟而定; X_{i2} 是第 i 个受感染个体的中心化的年龄; X_{i3} 是第 i 个受感染个体的中心化的感染前 CD4 百分率, 其值是由来自第 i 个个体的实际感染前 CD4 百分率的样本减其平均值来计算. 使用中心化的协变量 X_{i2} 和 X_{i3} 的优点是 $\beta_l(t)(l=0,\cdots,3)$ 有明确的生物学解释. 这里, $\beta_0(t)$ 表示基准 CD4 百分率, 且可以解释为对具有平均感染前 CD4 百分率和在感染下平均年龄的一个非吸烟者感染后在 t 年处平均 CD4 百分率, 于是, $\beta_1(t)$、$\beta_2(t)$ 和 $\beta_3(t)$ 在时间 t 基于感染后 CD4 百分率分别描述在感染下吸烟、年龄和感染前 CD4 百分率的影响.

我们使用标准 Gauss 核和 CV 带宽 $h_{\rm CV}=0.48$ 来计算 $\beta_r(t)(r=0,1,2,3)$ 的经验似然比和估计量. 为计算简单, 我们对 $f(t)$, $\gamma_{lr}(t)$, $b_l(t)$ 和 $v^2(t)$ 的估计量也使用标准 Gauss 核和 $h_{\rm CV}=0.48$. 我们取 $M=120$ 和 $c_1=3$ 并使用 RAE 和 NA 给出 $\beta_r(t)(r=0,1,2,3)$ 的 MELE 曲线和他们的相应 95% Bonferroni 类同时置信带. 为与 NA 比较我们选择 RAEL, 因为 Wu 等 (1998) 已经使用正态逼近来构造置信带. 图 5.4.2 报告了模拟结果.

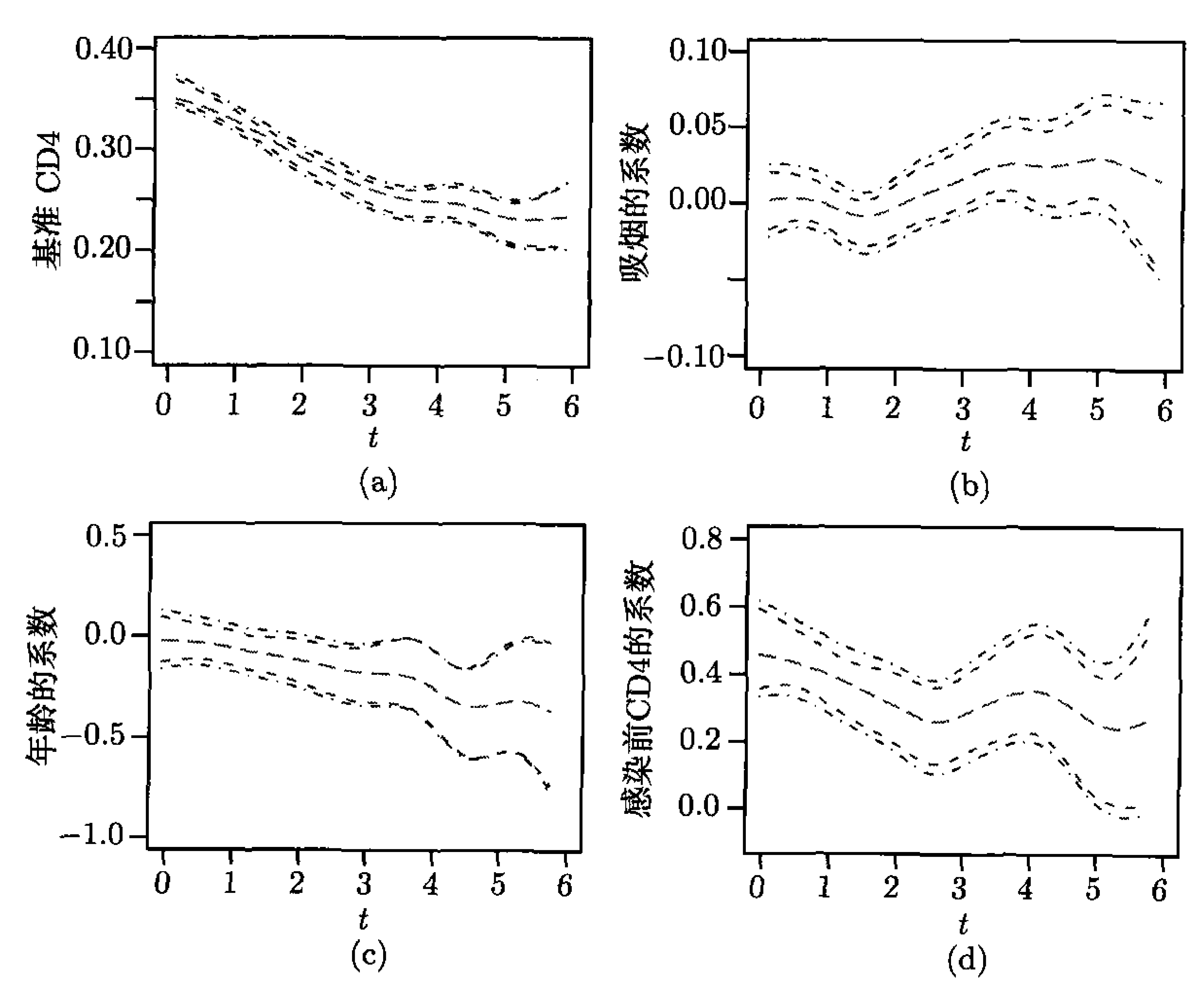


图 5.4.2 应用于艾滋病数据. 基于 CD4 细胞百分率, 基准 CD4 百分率和吸烟、年龄和感染前 CD4 百分率的估计的系数曲线. 实线为估计的系数曲线, 虚线和点线为基于 RAEL 和 NA 的 95% Bonferroni 类同时置信带. 图 (a) 是截距影响, (b) 是吸烟影响, (c) 是年龄影响, (d) 是感染前 CD4 影响

从图 5.4.2 我们发现总体的平均基准 CD4 百分率在 HIV 感染开始相当迅速地递减, 但感染四年以后递减的速度似乎慢了下来; 在 HIV 感染期间吸烟和年龄的确没有对感染后 CD4 百分率产生任何重大影响; 且感染前 CD4 百分率显现它与高感染后 CD4 百分率有十分肯定的联系. 这些发现与 Wu 和 Chiang(2000) 的核光滑方法所发现的结论基本一致. 因为在这里使用的吸烟变量是随时间变化的, 仅考虑"总是吸烟者" 和 "从未吸烟者". 所以图 5.4.2(b) 显示的曲线与 Wu 等 (1998) 的图 5.4.2(c) 稍有不同, 其中吸烟是时间相依的. 从图 5.4.2 我们也看到 RREL 显然优于 NA.

5.4.5 定理的证明

我们首先引入两个引理. 引理 5.4.1 提供了将在定理 5.4.1 的证明中使用的残差的上界, 引理 5.4.2 提供了 $\hat{b}_l(t_0)(l=0,\cdots,p)$ 的相合性.

引理5.4.1 设条件 (C5.3.2)~(C5.3.5) 成立, 且带宽满足 $Nh^5=O(1)$. 如果 $\beta(t_0)$ 是真参数, 则

$$\max_{1\leqslant i\leqslant n}||Z_i(\beta(t_0))||=o_P\big((Nh)^{1/2}\big).$$

证 $$U_i(t_0)=\sum_{j=1}^{m_i}X_i(t_{ij})\varepsilon_i(t_{ij})K\left(\frac{t_{ij}-t_0}{h}\right),$$

$$V_i(t_0)=\sum_{j=1}^{m_i}X_i(t_{ij})X_i^{\mathrm{T}}(t_{ij})[\beta(t_{ij})-\beta(t_0)]K\left(\frac{t_{ij}-t_0}{h}\right).$$

那么有 $Z_i(\beta(t_0))=U_i(t_0)+V_i(t_0)$. 由 Markov 不等式, 对任何 $\epsilon>0$ 和 $\delta>0$, 得

$$\begin{aligned}P\Big\{\max_{1\leqslant i\leqslant n}||U_i(t_0)||>\epsilon(Nh)^{1/2}\Big\}&\leqslant\sum_{i=1}^{n}E\big[||U_i(t_0)||^{2+\delta}\big]/[\epsilon(Nh)^{1/2}]^{2+\delta}\\&\leqslant c_1N^{-2\delta/5}+c_2N^{-(5\delta-2)/10}\left(N^{-6/5}\sum_{i=1}^{n}m_i^2\right)\to0,\end{aligned}$$

其中 c_1 和 c_2 是正的常数. 因此, 我们得到

$$\max_{1\leqslant i\leqslant n}||U_i(t_0)||=o_P\big((Nh)^{1/2}\big).$$

类似地, 可以得到

$$\max_{1\leqslant i\leqslant n}||V_i(t_0)||=o_P\big((Nh)^{1/2}\big),$$

这就证明了引理 5.4.1.

引理5.4.2 设定理 5.4.1 的条件满足, 则 $\hat{b}_l(t_0)\xrightarrow{P}b_l(t_0)$, $l=0,\cdots,p$.

证　记

$$\eta_{il}(t_0,t_{ij})=\sum_{r=0}^{p}X_{il}(t_{ij})X_{ir}(t_{ij})[\beta_r(t_{ij})-\beta_r(t_0)]K\left(\frac{t_{ij}-t_0}{h}\right),$$

$$\hat{\eta}_{il}(t_0,t_{ij})=\sum_{r=0}^{p}X_{il}(t_{ij})X_{ir}(t_{ij})[\hat{\beta}_r(t_{ij})-\hat{\beta}_r(t_0)]K\left(\frac{t_{ij}-t_0}{h}\right).$$

那么有

$$\begin{aligned}\hat{b}_l(t_0)-b_l(t_0)&=\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}[\hat{\eta}_{il}(t_0,t_{ij})-\eta_{il}(t_0,t_{ij})]\\&\quad+\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}[\eta_{il}(t_0,t_{ij})-(h/N)^{1/2}b_l(t_0)]\\&\equiv H_1(t_0)+H_2(t_0).\end{aligned}$$

由条件 (C5.3.1)~(C5.3.5) 和 Taylor 公式, 可知 $H_1(t_0)$ 的每一被加项可以由下式限制:

$$\sum_{r=0}^{p}X_{il}(t_{ij})X_{ir}(t_{ij})\{[\hat{\beta}_r'(t_0)-\beta_r'(t_0)](t_{ij}-t_0)+o_P(|t_{ij}-t_0|^2)\}K\left(\frac{t_{ij}-t_0}{h}\right).$$

使用定理 5.4.1 的条件可以证明

$$\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{r=0}^{p}X_{il}(t_{ij})X_{ir}(t_{ij})(t_{ij}-t_0)^{\nu}K\left(\frac{t_{ij}-t_0}{h}\right)=O_P(1),\quad \nu=1,2$$

和 $\hat{\beta}_r'(t_0)-\beta_r'(t_0)\xrightarrow{P}0$. 因此, $H_1(t_0)\xrightarrow{P}0$. 由条件 (C5.3.1)~(C5.3.5) 可得 $H_2(t_0)\xrightarrow{P}0$. 这就完成了引理 5.4.2 的证明.

有了前面的准备, 我们现在可以完成定理的证明.

定理 5.4.1 的证　由 (5.4.4) 式可以看到, 定理 5.4.1 的证明需要 $(Nh)^{-1/2}\sum_{i=1}^{n}Z_i(\beta(t_0))$ 的渐近正态性和 $\widetilde{D}(\beta(t))$ 的相合性. 首先证明

$$\tilde{\zeta}(\beta(t_0))\equiv\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}Z_i(\beta(t_0))\xrightarrow{D}N(0,\tilde{v}^2(t_0)\Gamma(t_0)),\tag{5.4.15}$$

其中 $\tilde{v}^2(t_0)=\sigma^2(t_0)f(t_0)\int K^2(u)du$. 类似于 Wu 等 (1998) 的引理 A1 的证明, 可以证明 $E[\tilde{\zeta}(\beta(t_0))]=o(1)$ 和 $\mathrm{cov}[\tilde{\zeta}(\beta(t_0))]=\tilde{v}^2(t_0)\Gamma(t_0)+o(1)$. 由引理 5.4.1 可知 $\tilde{\zeta}(\beta(t_0))$ 满足 Cramer-Wold 定理 (Serfling, 1980, §1.5.2 的定理) 和 Lindeberg 条件 (Serfling, 1980, §1.9.2 的定理). 因此 (5.4.15) 式成立.

由 Wu 等 (1998) 的证明可得

$$\frac{1}{Nh}\sum_{i=1}^{n} Z_i(\beta(t_0))Z_i^{\mathrm{T}}(\beta(t_0)) \xrightarrow{P} \tilde{v}^2(t_0)\Gamma(t_0). \tag{5.4.16}$$

由 (5.4.15) 和 (5.4.16) 式以及引理 5.4.1, 且使用文献 (Owen, 1990) 中 (2.14) 式同样的论证方法可以证明

$$\theta = O_P\big((Nh)^{-1/2}\big), \tag{5.4.17}$$

其中 θ 在 (5.4.3) 式中定义. 将 Taylor 公式应用到 (5.4.2) 式, 并调用 (5.4.15)~(5.4.17) 式和引理 5.4.1, 得

$$\mathcal{R}(\beta(t_0)) = 2\sum_{i=1}^{n}\left\{\theta^{\mathrm{T}} Z_i(\beta(t_0)) - \left[\theta^{\mathrm{T}} Z_i(\beta(t_0))\right]^2/2\right\} + o_P(1). \tag{5.4.18}$$

由 (5.4.3) 式可以推出

$$\begin{aligned}0 &= \sum_{i=1}^{n}\frac{Z_i(\beta(t_0))}{1+\theta^T Z_i(\beta(t_0))}\\ &= \sum_{i=1}^{n} Z_i(\beta(t_0)) - \sum_{i=1}^{n}\theta^{\mathrm{T}} Z_i(\beta(t_0))Z_i(\beta(t_0)) + \sum_{i=1}^{n}\frac{Z_i(\beta(t_0))[\theta^{\mathrm{T}} Z_i(\beta(t_0))]^2}{1+\theta^{\mathrm{T}} Z_i(\beta(t_0))}.\end{aligned}$$

应用 5.4.1 和 (5.4.15)~(5.4.17) 式可得

$$\sum_{i=1}^{n}[\theta^{\mathrm{T}} Z_i(\beta(t_0))]^2 = \sum_{i=1}^{n}\theta^T Z_i(\beta(t_0)) + o_P(1), \tag{5.4.19}$$

$$\theta = \left\{\sum_{i=1}^{n} Z_i(\beta(t_0))Z_i^T(\beta(t_0))\right\}^{-1}\sum_{i=1}^{n} Z_i(\beta(t_0)) + o_P\big((Nh)^{-1/2}\big). \tag{5.4.20}$$

将 (5.4.19) 和 (5.4.20) 式代入到 (5.4.18) 式可得 (5.4.4) 式. 结合 (5.4.4)、(5.4.15) 和 (5.4.16) 式即证明了定理 5.4.1.

定理 5.4.2 的证 由 (5.4.4) 和 (5.4.2) 式可得

$$\begin{aligned}\widetilde{\mathcal{R}}(\beta(t_0)) &= \left[\tilde{\zeta}(\beta(t_0)) - b(t_0)\right]^{\mathrm{T}}\tilde{D}^{-1}(\beta(t_0))\left[\tilde{\zeta}(\beta(t_0)) - b(t_0)\right]\\ &\quad + \left[\hat{b}(t_0) - b(t_0)\right]^{\mathrm{T}}\tilde{D}^{-1}(\beta(t_0))\left[\hat{b}(t_0) + b(t_0) - 2\tilde{\zeta}(\beta(t_0))\right] + o_P(1)\\ &\equiv L_1(t_0) + L_2(t_0) + o_P(1).\end{aligned}$$

类似于 (Wu et al., 1998) 中 (A.12) 式的证明, 可以证得

$$\widetilde{D}(\beta(t_0)) \xrightarrow{P} v^2(t_0)\Gamma(t_0),$$

$$\tilde{\zeta}(\beta(t_0)) - b(t_0) \xrightarrow{D} N(0, v^2(t_0)\Gamma(t_0)).$$

结合引理 5.4.2 即证得 $L_1(t_0) \xrightarrow{D} \chi^2_{p+1}$ 和 $L_2(t_0) \xrightarrow{P} 0$. 这就完成了定理 5.4.2 的证明.

定理 5.4.3 的证明类似于定理 5.4.1 的证明, 因此这里略去它的证明.

命题 5.4.1 的证　仅证明 $\hat{v}^2(t_0) \xrightarrow{P} v^2(t_0)$, 其他各式的证明由引理 5.4.2 和 Wu 等 (1998) 给出. 记

$$\psi_i(t_0) = \sum_{j=1}^{m_i} \{X_i^{\mathrm{T}}(t_{ij})[\beta(t_{ij}) - \beta(t_0)] + \varepsilon_i(t_{ij})\} K\left(\frac{t_{ij} - t_0}{h}\right),$$

$$\phi_i(t_0) = \sum_{j=1}^{m_i} X_i^{\mathrm{T}}(t_{ij})[\beta(t_0) - \hat{\beta}(t_0)] K\left(\frac{t_{ij} - t_0}{h}\right).$$

由 (5.4.6) 式可得 $\hat{v}^2(t_0) = (Nh)^{-1}\sum_{i=1}^{n}\left[\psi_i(t_0) + \phi_i(t_0)\right]^2$. 因此,

$$\begin{aligned}\hat{v}^2(t_0) - v^2(t_0) &= \left\{\frac{1}{Nh}\sum_{i=1}^{n}\psi_i^2(t_0) - v^2(t_0)\right\} \\ &\quad + \frac{1}{Nh}\sum_{i=1}^{n}\phi_i^2(t_0) + \frac{2}{Nh}\sum_{i=1}^{n}\psi_i(t_0)\phi_i(t_0) \\ &\equiv J_1(t_0) + J_2(t_0) + 2J_3(t_0).\end{aligned}$$

由命题 5.4.1 的条件可得

$$\begin{aligned}E[J_1^2(t_0)] &\leqslant 2E\left\{\frac{1}{Nh}\sum_{i=1}^{n}[\psi_i(t_0) - E(\psi_i(t_0))]\right\}^2 \\ &\quad + 2\left\{\frac{1}{Nh}\sum_{i=1}^{n}E[\psi_i(t_0)] - v^2(t_0)\right\}^2 \\ &\leqslant 4(Nh)^{-2}\sum_{i=1}^{n}E[\psi_i^2(t_0)] + o(1) \\ &\leqslant 4N^{-2}h^{-1}\sum_{i=1}^{n}m_i^2 + o(1) = o(1).\end{aligned}$$

从而, $J_1(t_0) = o_P(1)$. 由定理 5.4.1 和条件 (C5.3.1)~(C5.4.4) 可得

$$\hat{\beta}(t_0) - \beta(t_0) = O_P\big((Nh)^{-1/2}\big),$$

$$\frac{1}{Nh}\sum_{i=1}^{n}\left[\sum_{j=1}^{m_i}||X_i(t_{ij})||K\left(\frac{t_{ij} - t_0}{h}\right)\right]^2 = O_P\big(N^{1/5}\big).$$

因此, $J_2(t_0)=o_P(1)$. 由 Cauchy-Schwarz 不等式可得

$$J_3(t_0)\leqslant\left\{\frac{1}{Nh}\sum_{i=1}^{n}\psi_i^2(t_0)\right\}^{1/2}\left\{\frac{1}{Nh}\sum_{i=1}^{n}\phi_i^2(t_0)\right\}^{1/2}=o_P(1).$$

结合 $J_1(t_0)=o_P(1)$ 和 $J_2(t_0)=o_P(1)$ 即证明了 $\hat{v}^2(t_0)\xrightarrow{P}v^2(t_0)$.

参考文献

张日权, 卢一强. 2004. 变系数模型. 北京: 科学出版社

Akaike H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19: 716~723

Arnold S F. 1981. The Theory of Linear Models and Multivariate Analysis. New York: John Wiley & Sons

Boor C. 1978. A practical guide to splines. New York: Springer

Cai Z W. 2002. Two-step likelihood estimation procedure for varying-coefficient models. Journal of Multivariate Analysis, 82(1): 189~209

Cai Z W. 2003. Local quasi-likelihood approach to varying-coefficient discrete-valued time series models. Journal of Nonparametric Statistics 15(6): 693~711

Cai Z W, Fan J, Li Y Z. 2000a. Efficient estimation and inferences for varying-coefficient models. Journal of the American Statistical Association, 95(451): 888~902

Cai Z W, Fan J, Yao Q. 2000b. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association, 95(451): 941~956

Chiang C T, Rice J A, Wu C O. 2001. Smoothing spline estimation for varying coefficient models with repeatedly measure dependent variables. Journal of the American Statistical Association, 96: 605~619

Cleveland W H, Grosse E, Shyu W M. 1991. Local regression models. Statistical Models (eds J M Chambers and T J Hastie). pp. 309~376, Pacific Grove: Wadsworth and Brooks

DeVore R A, Lorentz G G.1993. Constructive approximation. Berlin: Springer-Verlag

Eggermont P P B, Eubank R L, LaRiccia V N. 2010. Convergence rates for smoothing spline estimators in varying coefficient models. Journal of Statistical Planning and Inference, 140(2): 369~381

Eubank R L, Huang C F, Maldonado Y M, et al. 2004. Smoothing spline estimation in varying-coefficient models. Journal of the Royal Statistical Society Series B, 66: 653~667

Fan J. 1992. Design-adaptive nonparametric regression, Journal of the American Statistical Association, 87: 998~1004

Fan J, Huang T. 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli, 11(6): 1031~1057

Fan J, Yao Q, Cai Z W. 2003. Adaptive varying-coefficient linear models. Journal of the Royal Statistical Society, Series B, 65: 57~80

Fan J, Zhang W Y. 1999. Statistical estimation in varying-coefficient models. Annals of Statistics, 27 (5): 1491~1518

Fan J, Zhang W 2000a. Simultanenous confidence bands and hypothsis testing in vary-coefficient models. Scandinavian Journal of Statistics, 27: 715~731

Fan J, Zhang J. 2000b. Two-step estimation of functional linear models with applications to longitudinal data. Journal of the Royal Statistical Society Series B, 62: 303~322

Galindo C D, Liang H, Kauermann G, et al. 2001. Bootstrap confidence intervals for local likelihood, local estimating equations and varying coefficient models. Statistica Sinica, 11(1): 121~134

Gelfand A E, Kim H J, Sirmans C F, et al. 2003. Spatial modeling with spatially varying coefficient processes. Journal of the American Statistical Association, 98(462): 387~396

Guo W S. 2004. Functional data analysis in longitudinal settings using smoothing splines, Statistical Methods in Medical Research, 13(1): 49~62

Hall P, Titterington D M. 1988. On confidence bands in nonparametric density estimation and regression. Journal of Multivariate Analysis, 27: 228~254

Härdle W. 1989. Applied Nonparametric Regression. Cambridge: Cambridge University Press

Hastie T J, Tibshirani R J. 1990. Generalized additive models. London: Chapman & Hall

Hastie T J, Tibshirani R J. 1993. Varying-coefficient models. Journal of the Royal Statistical Society Series B, 55: 757~796

Härdle W, Liang H, Gao J T. 2000. Partially linear models. Heidelberg: Spring Physica

Heim S, Fahrmeir L, Eilers P H C, et al. 2007. 3D space-varying coefficient models with application to diffusion tensor imaging. Computational Statistics & Data Analysis, 51(12): 6212~6228

Hogan J W, Lin X H, Herman B. 2004. Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. Biometrics, 60(4): 854~864

Homg Z P, Lian H. 2012. Time-varying coefficient estimation in differential equation models with noisy time-varying covariates. Journal of Multivariate Analysis, 103(1): 58~67

Honda T. 2004. Quantile regression in varying coefficient models. Journal of Statistical Planning and Inference 121(1): 113~125

Hoover D R, Rice J A, Wu C O, et al. 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika, 85(4): 809~822

Hu X M, Wang Z Z, Zhao Z W. 2009. Empirical likelihood for semiparametric varying-

coefficient partially linear errors-in-variables models. Statistics & Probability Letters, 79(8): 1044~1052

Huang J Z. 1998. Projection estimation in multiple regression with applications to functional ANOVA models. Annals Statistics, 26: 242~272

Huang J Z, Shen H P. 2004. Functional Coefficient Regression Models for Non-linear Time Series: A Polynomial Spline Approach. Scandinavian Journal of Statistics, 31(4): 515~534

Huang J Z, Wu C O, Zhou L. 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika, 89(1): 111~128

Huang J Z, Wu C O, Zhou L. 2004. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statistica Sinica, 14(3): 763~788

Hurvich C M, Tsai C L. 1989. Regression and time series model selection in small samples. Biometrika, 76: 297~307

Ip W C, Wong H N, Zhang R Q. 2007. Generalized likelihood ratio test for varying-coefficient models with different smoothing variables. Computational Statistics & Data Analysis, 51(9): 4543~4561

James G M, Silverman B W. 2005. Functional adaptive model estimation. Journal of the American Statistical Association, 100(470): 565~576

Kauermann G. 2000. Modeling longitudinal data with ordinal response by varying coefficients. Biometrics, 56(3): 692~698

Kauermann G, Tutz G, Bruderl J. 2005. The survival of newly founded firms: a case-study into varying-coefficient models. Journal of the Royal Statistical Society, Series A-Statistics in Society, 168: 145~158

Kim M. 2007. Quantile regression with varying coefficients. The Annals of Statistics, 35: 92~108

Knafl G, Sacks J, Ylvisaker D. 1985. Confidence bands for regression functions. Journal of the American Statistical Association, 80: 683~691

Krafty R T, Gimotty P A, Holtz D, et al. 2008. Varying coefficient model with unknown within-subject covariance for analysis of tumor growth curves. Biometrics, 64(4): 1023~1031

Kuruwita C N, Kulasekera K B, Gallagher C M. 2011. Generalized varying coefficient models with unknown link function. Biometrika, 98(3): 701~710

Lee D, Shaddick G. 2007. Time-varying coefficient models for the analysis of air pollution and health outcome data. Biometrics, 63(4): 1253~1261

Leng C L. 2009. A simple approach for varying-coefficient model selection. Journal of Statistical Planning and Inference, 139(7): 2138~2146

Li N, Xu X Z, Liu X H. 2011. Testing the constancy in varying-coefficient regression models. Metrika, 74(3): 409~438

Li Z Q, Xue L G. 2010. Mixtures of semiparametric varying coefficient models for longitudinal data with nonignorable dropout. Acta Mathematicae Applicatae Sinica-English Series, 26(1): 125~132

Liang H, Wu H L, Carroll R J. 2003. The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error. Biostatistics, 4(2): 297~312

Lin D Y, Ying Z. 2001. Semiparametric and nonparametric regression analysis of longitudinal data. Journal of the American Statistical Association, 96(453): 116~119

Lu Y Q, Mao S S. 2004. Local asymptotics for b-spline estimators of the varying coefficient model. Communications in Statistics-Theory and Methods, 33(5): 1119~1138

Lu Y Q, Zhang R Q. 2009. Smoothing spline estimation of generalised varying-coefficient mixed model. Journal of Nonparametric Statistics, 21(7): 815~825

Lv Y Z, Zhang R Q, Huang Z S. 2011. Nonparametric estimation of varying coefficient error-in-variable models with validation sampling. Journal of statistical planning and inference, 141(10): 3323~3344

Noh H S, Park B U. 2010. Sparse varying coefficient models for longitudinal data. Statistica Sinica, 20(3): 1183~1202

Owen A B. 1988. Empirical likelihood ratio confidence intervals for a single function. Biometrika, 75: 237~249

Owen A B. 1990. Empirical likelihood ratio confidence regions. The Annals of Statistics, 18: 90~120

Park B U, Hwang J H, Park M S. 2011. Testing in nonparametric varying coefficient additive models. Statistica Sinica, 21(2): 749~778

Ramsay J O, Silverman B W. 1997. Functional Data Analysis. New York: Springer-Verlag

Rice J A, Silverman B W. 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society, Series B, 53: 233~243

Schumaker L. 1981. Spline Functions: Basic Theory. New York: John Wiley & Sons

Schwarz G. 1978. Estimating the dimension of a model. The Annals of Statistics, 6: 461~464

Sentuerk D, Nguyen D V. 2011. Varying coefficient models for sparse noise-contaminated longitudinal data. Statistica sinica, 21(4): 1831~1856

Sentürk D, Müller H G. 2005. Covariate adjusted correlation analysis via varying coefficient models. Scandinavian Journal of Statistics, 32(3): 365~383

Sentürk D, Müller H G. 2008. Generalized varying coefficient models for longitudinal data. Biometrika, 95(3): 653~666

Serfling R. 1980. Approximation Theorems of Mathematical Statistics. New York: Wiley

Shumway R H. 1988. Applied Statistical Time Series Analysis. Prentice-Hall

Stone C J. 1982. Optimal global rates of convergence for nonparametric regression. Annals of Statistics. 10: 1348~1360

Su L, Hogan J W. 2010. Varying-coefficient models for longitudinal processes with continuous-time informative dropout. Biostatistics, 11(1): 93~110

Tang Q G, Cheng L S. 2009a. B-spline estimation for varying coefficient regression with spatial data. Science in China, Series A-Mathematics, 52(11): 2321~2340

Tang Q G, Cheng L S. 2009b. Asymptotic Normality of M-Estimators for Varying Coefficient Models with Longitudinal Data. Communications in Statistics-Theory and Methods, 38(9): 1422~1440

Tang Q G, Wang J D. 2005a. L_1-estimation for varying coefficient models. Statistics, 39(5): 389~404

Tang Q G, Wang J D. 2005b. One-step estimation for varying coefficient models. Science in China, Series a-Mathematics, 48(2): 198~213

Tjϕstheim D, Auestad B H. 1994. Nonparametric identification of nonlinear time series: projections. Journal of the American Statistical Association, 89: 1398~1409

Wang H S, Xia Y C. 2009. Shrinkage estimation of the varying coefficient model. Journal of the American Statistical Association, 104(486): 747~757

Wang L, Kai B, Li R Z. 2009. Local rank inference for varying coefficient models. Journal of the American Statistical Association, 104(488): 1631~1645

Wang L F, Li H Z, Huang J H Z. 2008. Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. Journal of the American Statistical Association, 103(484): 1556~1569

Wang Q H, Zhang R Q. 2009. Statistical estimation in varying coefficient models with surrogate data and validation sampling. Journal of Multivariate Analysis, 100(10): 2389~2405

Wei C H. 2011. Estimation in varying-coefficient errors-in-variables models with missing response variables. Communications in Statistics-Simulation and Computation, 40(3): 383~393

Wei F R, Huang J, Li H Z. 2011. Variable selection and estimation in high-dimensional varying-coefficient models. Statistica sinica, 21(4): 1515~154

Wong H, Guo S J, Chen M, et al. 2009. On locally weighted estimation and hypothesis testing of varying-coefficient models with missing covariates. Journal of Statistical Planning and Inference, 139(9): 2933~2951

Wu C O, Chiang C T. 2000. Kernel smoothing on varying coefficient models with longitudinal dependent variable. Statistica Sinica, 10: 433~456

Wu C O, Chiang C T, Hoover D R. 1998. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. Journal of the American Statistical Association, 93: 1388~1403

Wu H L, Liang H. 2004. Backfitting random varying-coefficient models with time-dependent smoothing covariates. Scandinavian Journal of Statistics, 31(1): 3~19

Wu C O, Yu K F, Chiang C T. 2000. A two-step smoothing method for varying-coefficient models with repeated measurements. Annals of the Institute of Statistical Mathematics, 52(3): 519~543

Xia Y C, Li W K. 1999. On the estimation and testing of functional-coefficient linear models. Statistica Sinica, 9: 735~757

Xu W L, Zhu L X. 2008. Goodness-of-fit testing for varying-coefficient models. Metrika, 68(2): 129~146

Xue L G, Zhu L X. 2007. Empirical likelihood for a varying coefficient model with longitudinal data. Journal of the American Statistical Association, 102(478): 642~654

Zhang D W. 2004. Generalized linear mixed models with varying coefficients for longitudinal data. Biometrics, 60(1): 8~15

Zhang W Y. 2002. Local polynomial fitting in semivarying coefficient model. Journal of Multivariate Analysis, 82: 166~188

Zhang W Y, Lee S Y. 2000. Variable bandwidth selection in vary-coefficient models. Journal of Multivariate Analysis, 74: 116~134

Zhang W Y, Peng H. 2010. Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. Journal of Multivariate Analysis, 101(7): 1656~1680

Zhao P X, Xue L G. 2011. Variable selection for varying coefficient models with measurement errors. Metrika, 74(2): 231~245

第 6 章 部分线性变系数模型

6.1 引 言

设 Y 是响应变量, (U, X, Z) 是联合协变量, 那么部分线性变系数模型具有如下形式:

$$Y = \beta_0^{\mathrm{T}} Z + \boldsymbol{a}^{\mathrm{T}}(U) X + \varepsilon, \tag{6.1.1}$$

其中 $\boldsymbol{a}(\cdot) = (a_1(\cdot), \cdots, a_p(\cdot))^{\mathrm{T}}$ 是未知函数向量, β_0 是 q 维未知参数向量, ε 是随机误差, 且满足 $E(\varepsilon) = 0$ 和 $\mathrm{var}(\varepsilon) = \sigma^2$. 由于维数灾祸, 为方便起见, 假定 U 是一维协变量. 模型 (6.1.1) 允许 U 与 X 之间以某种方式发生交互作用, 即不同的水平 U 联系不同的线性模型. 这就允许我们考察协变量 X 的影响在不同的变量 U 的水平上的变化程度. 当 $p = 1$ 且 $X = 1$ 时, 模型 (6.1.1) 变成部分线性模型. 因此, 这个模型比单纯的变系数模型更加灵活, 也具有降维的能力.

许多学者对部分线性变系数模型进行了深入研究, 得到了一些理想的研究成果. Zhang(2002) 通过对函数系数用局部多项式拟合, 构造了模型中参数和非参数分量的估计. Xia 等 (2004) 提出了一个新的估计方法, 构造了模型参数的一个有效估计, 建立了估计量的渐近性质, 并发展了模型选择方法. Fan 和 Huang(2005) 提出了一个估计参数分量的 profile 最小二乘方法, 研究了估计量的渐近正态性, 并引入了 profile 似然比检验和 Wald 类统计量. Ahmad 等 (2005) 提出了广义级数方法来估计模型中的兴趣参数, 建立了估计量的相合性和渐近正态性, 并在误差方差是条件同方差的情况下证明了这个估计量是半参数有效的. Li 和 Liang(2008) 将 SCAD 惩罚应用到了广义部分线性变系数模型的变量选择, 对参数部分, 采用非凸惩罚似然提出了一类变量选择方法, 证明了所得的惩罚估计具有 oracle 性质; 对非参数部分, 提出了半参数广义似然比检验, 通过检验选出非参数部分中有效变量, 同时研究了在零假设下所提出的检验统计量的渐近性质. Lam 和 Fan(2008) 考虑当参数分量的维数随着样本大小趋向于无穷大时, 模型中参数和非参数分量的估计的渐近性质. Wang 等 (2009) 利用假设检验方法研究了部分线性变系数分位数回归的模型选择问题, 并对基函数逼近和惩罚方法的变量选择过程进行了讨论. Zhao 和 Xue(2009a) 结合基函数逼近和惩罚最小二乘, 提出了一个变量选择方法, 并且研究了该方法的统计性质. 另外, 利用局部二次逼近方法, 给出了计算正则估计的一个迭代算法. 关于部分线性变系数模型的其他研究工作可参见文献 (Zhou and You,

2004; You and Chen, 2006; You and Zhou, 2006; Hu et al., 2009; Huang and Zhang, 2009; Li et al., 2009; Li, Chen and Lin, 2011; Lia and Palta, 2009; Zhao and Xue, 2009b, 2010a, 2010b; Zhou and Liang, 2009; Zhou et al., 2009; Huang et al., 2010; Hu and Cui, 2010; Yang and Li, 2010; Huang, 2011; Kai et al., 2011; Li, Feng and Peng, 2011; Li, Xue and Lian, 2011; Wang et al., 2011; Zhang et al., 2011) 等.

6.2 估计方法

6.2.1 profile 最小二乘估计

1. *估计方法*

有许多估计未知参数 β 和系数函数 $\boldsymbol{a}(\cdot)$ 的方法. 一个很有用的方法是 profile 最小二乘方法, 可以证明它是一个半参数有效方法. 当 $\varepsilon \sim N(0,\sigma^2)$ 时, 这个方法就成为 profile 似然方法.

假设样本 $\{(U_i,X_i,Z_i,Y_i), 1\leqslant i\leqslant n\}$ 是来自模型 (6.1.1) 的 iid 样本, 则有

$$Y_i=\beta_0^{\mathrm{T}}Z_i+\boldsymbol{a}^{\mathrm{T}}(U_i)X_i+\varepsilon_i,\quad i=1,\cdots,n, \tag{6.2.1}$$

其中 $X_i=(X_{i1},\cdots,X_{ip})^{\mathrm{T}}$, $Z_i=(Z_{i1},\cdots,Z_{iq})^{\mathrm{T}}$. 对给定的 β_0, 可将模型 (6.2.1) 写作

$$Y_i^*=\sum_{j=1}^{p}a_j^{\mathrm{T}}(U_i)X_{ij}+\varepsilon_i,\quad i=1,\cdots,n, \tag{6.2.2}$$

其中 $Y_i^*=Y_i-\beta_0^{\mathrm{T}}Z_i$. 这就将部分线性变系数模型 (6.2.1) 转换成了变系数模型 (6.2.2). 下面利用局部线性回归技术来估计模型 (6.2.2) 中的系数函数 $\{a_j(\cdot), j=1,\cdots,p\}$. 对于给定的点 u_0, 在 u_0 的一个邻域内用线性函数 $a_j(u)\approx a_j+b_j(u-u_0)$ 局部地逼近 $a_j(\cdot)$. 这就导致下列局部加权最小二乘问题: 极小化

$$\sum_{i=1}^{n}\left\{Y_i^*-\sum_{j=1}^{p}[a_j+b_j(U_i-u_0)]X_{ij}\right\}^2 K\left(\frac{U_i-u_0}{h}\right), \tag{6.2.3}$$

来求 $\{a_j(\cdot), j=1,\cdots,,p\}$, 其中 $K(\cdot)$ 是核函数, $h=h_n>0$ 是带宽序列. 为使用矩阵记号, 记 $\boldsymbol{Y}=(Y_1,\cdots,Y_n)^{\mathrm{T}}$, $\boldsymbol{Z}=(Z_1,\cdots,Z_n)^{\mathrm{T}}$, $W_u=\mathrm{diag}\{K((U_1-u)/h),\cdots,K((U_p-u)/h)\}$,

$$D_u=\begin{pmatrix} X_1^{\mathrm{T}} & X_1^{\mathrm{T}}(U_1-u)\\ \vdots & \vdots\\ X_n^{\mathrm{T}} & X_n^{\mathrm{T}}(U_n-u)\end{pmatrix}.$$

由最小二乘理论可求得极小化问题 (6.2.3) 的解为

$$\hat{a}_j(u;\beta_0) = \boldsymbol{e}_{j,2p}^{\mathrm{T}}(D_u^{\mathrm{T}}W_uD_u^{\mathrm{T}})^{-1}D_u^{\mathrm{T}}W_u(\boldsymbol{Y}-\boldsymbol{Z}\beta_0), \tag{6.2.4}$$

其中 $\boldsymbol{e}_{j,2p}$ 是 $2p\times 1$ 单位向量, 其第 j 个分量是 1. 记 $\hat{\boldsymbol{a}}(\cdot;\beta_0) = (\hat{a}_1(\cdot;\beta_0),\cdots,\hat{a}_p(\cdot;\beta_0))^{\mathrm{T}}$. 用 $\hat{\boldsymbol{a}}(U_i;\beta_0)$ 代替 (6.2.1) 式中的 $\boldsymbol{a}(U_i)$ 可得

$$Y_i = \beta_0^{\mathrm{T}}Z_i + \hat{\boldsymbol{a}}^{\mathrm{T}}(U_i;\beta_0)X_i + \epsilon_i, \quad i=1,\cdots,n.$$

用 S 记局部线性回归的光滑矩阵, 即

$$S = \begin{pmatrix} [X_1^{\mathrm{T}}\ \ 0](D_{U_1}^{\mathrm{T}}W_{U_1}D_{U_1})^{-1}D_{U_1}^{\mathrm{T}}W_{U_1} \\ \vdots \\ [X_n^{\mathrm{T}}\ \ 0](D_{U_n}^{\mathrm{T}}W_{U_n}D_{U_n})^{-1}D_{U_n}^{\mathrm{T}}W_{U_n}. \end{pmatrix},$$

则有

$$(I-S)\boldsymbol{Y} = (I-S)\boldsymbol{Z}\beta_0 + \boldsymbol{\epsilon},$$

其中 $\boldsymbol{\epsilon} = (\varepsilon_1,\cdots,\varepsilon_n)^{\mathrm{T}}$. 将线性模型的最小二乘理论应用到上式可得

$$\hat{\beta}_{\mathrm{P}} = [\boldsymbol{Z}^{\mathrm{T}}(I-S)^{\mathrm{T}}(I-S)\boldsymbol{Z}]^{-1}\boldsymbol{Z}^{\mathrm{T}}(I-S)^{\mathrm{T}}(I-S)\boldsymbol{Y}. \tag{6.2.5}$$

称 $\hat{\beta}_{\mathrm{P}}$ 为 β_0 的最小二乘估计.

2. 主要结果

首先给出一组条件.

(C6.2.1) 随机变量 U 具有有界支撑 $\mathcal{U}$, 其密度函数 $f(\cdot)$ 是 Lipschitz 连续的且在 $\mathcal{U}$ 上有界远离 0.

(C6.2.2) 对每个 $u\in\mathcal{U}$, 矩阵 $E(XX^{\mathrm{T}}|U=u)$ 是正定的, 且 $E(XX^{\mathrm{T}}|U=u)$ 和 $E(XX^{\mathrm{T}}|U=u)^{-1}$ 以及 $E(XZ^{\mathrm{T}}|U=u)$ 都是 Lipschitz 连续的.

(C6.2.3) 存在 $s>0$ 使得 $E\|X\|^{2s}<\infty$, $E\|Z\|^{2s}<\infty$, 且对某个 $\epsilon<2-s^{-1}$ 使得 $n^{2\epsilon-1}h\to\infty$.

(C6.2.4) $\{a_j(u), j=1,\cdots,p\}$ 在 $\mathcal{U}$ 上有连续的 2 阶导数.

(C6.2.5) 核 $K(\cdot)$ 是对称的密度函数且具有有界支撑.

(C6.2.6) $nh^8\to 0$, $nh^2/\log^2 n\to\infty$.

下述定理给出了估计量 $\hat{\beta}_{\mathrm{P}}$ 的渐近分布, 其证明可参见文献 (Fan and Huang, 2005).

定理6.2.1 设条件 (C6.2.1)~(C6.2.5) 成立, 则 $\sqrt{n}(\hat{\beta}_{\mathrm{P}}-\beta_0)\xrightarrow{D}N(0,\Sigma)$, 其中

$$\Sigma = \sigma^2\{E(ZZ^{\mathrm{T}}) - E[E(ZX^{\mathrm{T}}|U)E(XX^{\mathrm{T}}|U)^{-1}E(XZ^{\mathrm{T}}|U)]\}^{-1}.$$

考虑 $p=1$ 且 $X\equiv 1$ 的情况, 此时模型 (6.1.1) 为部分线性模型, 那么

$$E(ZZ^{\mathrm{T}})-E[E(Z|U)E(Z^{\mathrm{T}}|U)]=E[\mathrm{var}(Z|U)].$$

事实上, 他们证明了 Σ 是半参数信息界 (Bickel and Kwon, 2001). 该结果当然也适合更一般的部分线性变系数模型 (Chamberlain, 1992). 因此, profile 似然估计是半参数有效的.

Fan 和 Huang(2005) 对参数分量的检验问题提出了 profile 似然比检验, 并得到了所提出的检验统计量在零假设下服从 χ^2 分布.

6.2.2　一般序列估计

在 6.2.1 小节中. Fan 和 Huang(2005) 使用核 profile 最小二乘方法来估计变系数部分线性模型中的兴趣参数, 并证明了他们的方法在同方差误差情况下导致了 β_0 的有效估计. 然而, 当误差是条件异方差时, 要使用核方法得到 β_0 的有效估计有更大的困难. 序列估计作为估计未知条件均值回归函数的最好逼近函数有好的定义内涵, 即使模型是错误指定时也是如此. 使用一般序列估计所付出的代价是: 要在最优光滑下 (即平衡平方偏差和方差项) 建立非参数分量估计的渐近正态性是困难的. 因此, 序列方法将被认为弥补部分线性变系数模型中核估计方法的不足.

1. *估计方法*

下面利用序列估计方法来估计模型 (6.1.1) 中的参数分量 β_0 和非参数分量 $a(\cdot)$. 假定随机误差 ε_i 满足 $E(\varepsilon_i|U_i,X_i,Z_i)=0$. 对 $j=1,\cdots,p$, 用 k_j 个基函数的线性组合 $\boldsymbol{g}_{j,k_j}(u)\alpha_{j,k_j}$ 逼近变系数函数 $a_j(u)$, 其中 $\boldsymbol{g}_{j,k_j}(u)=(g_{j1}(u),\cdots,g_{jk_j}(u))^{\mathrm{T}}$ 是 $k_j\times 1$ 基函数向量, $\alpha_{j,k_j}=(\alpha_{j1},\cdots,\alpha_{jk_j})^{\mathrm{T}}$ 是 $k_j\times 1$ 未知参数向量. 逼近函数 $g_{j,k_j}(u)$ 有如下特性: 正因为 k_j 可增大, 从而存在 $g_{j,k_j}(u)$ 的线性组合可很好地逼近任何光滑函数 $a_j(u)$, 其逼近的均方误差可以任意小.

定义 $K\times 1$ 矩阵 $\boldsymbol{g}_K(U_i,X_i)=(\boldsymbol{g}_{1,k_1}^{\mathrm{T}}(U_i)X_{i1},\cdots,\boldsymbol{g}_{p,k_p}^{\mathrm{T}}(U_i)X_{ip})^{\mathrm{T}}$ 和 $\alpha=(\alpha_{1,k_1}^{\mathrm{T}},\cdots,\alpha_{p,k_p}^{\mathrm{T}})^{\mathrm{T}}$, 其中 $K=\sum\limits_{j=1}^{p}k_j$. 从而, 使用 K 个函数 $\boldsymbol{g}_K(U_i,X_i)$ 的线性组合来逼近 $a^{\mathrm{T}}(U_i)X_i$. 因此, 可将 (6.2.1) 重写为

$$\begin{aligned}Y_i&=\beta_0^{\mathrm{T}}Z_i+\boldsymbol{g}_K^{\mathrm{T}}(U_i,X_i)\alpha+[a^{\mathrm{T}}(U_i)X_i-\boldsymbol{g}_K^{\mathrm{T}}(U_i,X_i)\alpha]+\varepsilon_i\\&=\beta_0^{\mathrm{T}}Z_i+\boldsymbol{g}_K^{\mathrm{T}}(U_i,X_i)\alpha+\mathrm{error}_i,\end{aligned}\tag{6.2.6}$$

其中误差的定义是显然的.

记 $A=(\boldsymbol{a}^{\mathrm{T}}(U_1)X_1,\cdots,\boldsymbol{a}^{\mathrm{T}}(U_n)X_n)^{\mathrm{T}}$, $G=(\boldsymbol{g}_K(U_1,X_1),\cdots,\boldsymbol{g}_K(U_n,X_n))^{\mathrm{T}}$,

$\boldsymbol{Y} = (Y_1, \cdots, Y_n)^{\mathrm{T}}$, $\boldsymbol{Z} = (Z_1, \cdots, Z_n)^{\mathrm{T}}$. 模型 (6.2.6) 可用矩阵形式写为

$$\boldsymbol{Y} = \boldsymbol{Z}\beta_0 + G\alpha + \text{error}. \tag{6.2.7}$$

用 $\hat{\alpha}$ 和 $\hat{\beta}_{\mathrm{GS}}$ 分别记 (6.2.7) 式中由 $\boldsymbol{Y}$ 关于 (G, Z) 的回归而得到的 α 和 β_0 的最小二乘估计. 那么, 可以用 $\hat{a}_j(u) \stackrel{\text{def}}{=} \boldsymbol{g}_{j,k_j}^{\mathrm{T}}(u)\hat{\alpha}_{j,k_j}(j = 1, \cdots, p)$ 估计 $a_j(u)$.

下面导出 $\hat{\alpha}$ 和 $\hat{\beta}_{\mathrm{GS}}$ 的具体表达式. (6.2.1) 式可以写作如下矩阵形式:

$$\boldsymbol{Y} = \boldsymbol{Z}\beta_0 + A + \varepsilon, \tag{6.2.8}$$

其中 $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^{\mathrm{T}}$. 记 $M = G(G^{\mathrm{T}}G)^{-}G^{\mathrm{T}}$, 这里 H^{-} 表示 H 的任何对称广义逆. 对任一 $n \times m$ 矩阵 B, 定义 $\widetilde{B} = MB$. 那么, 用 M 左乘 (6.2.8) 式两边可得

$$\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{Z}}\beta_0 + \tilde{A} + \tilde{\varepsilon}, \tag{6.2.9}$$

(6.2.8) 和 (6.2.9) 式左右两边分别相减可得

$$\boldsymbol{Y} - \widetilde{\boldsymbol{Y}} = (\boldsymbol{Z} - \widetilde{\boldsymbol{Z}})\beta_0 + (A - \tilde{A}) + \varepsilon - \tilde{\varepsilon}, \tag{6.2.10}$$

把 (6.2.10) 式看作 $\boldsymbol{Y} - \widetilde{\boldsymbol{Y}}$ 关于 $(\boldsymbol{Z} - \widetilde{\boldsymbol{Z}})$ 的线性回归, 并由最小二乘估计法可得到 $\hat{\beta}_{\mathrm{GS}}$, 即

$$\hat{\beta}_{\mathrm{GS}} = (\boldsymbol{Z} - \widetilde{\boldsymbol{Z}})^{\mathrm{T}}(\boldsymbol{Z} - \widetilde{\boldsymbol{Z}})^{-}(\boldsymbol{Z} - \widetilde{\boldsymbol{Z}})^{\mathrm{T}}(\boldsymbol{Y} - \widetilde{\boldsymbol{Y}})$$

用 $\hat{\beta}_{\mathrm{GS}}$ 代替 (6.2.7) 式中的 β_0, 并由最小二乘估计法可得到 $\hat{\alpha}$, 即

$$\hat{\alpha} = (G^{\mathrm{T}}G)^{-}G^{\mathrm{T}}(\boldsymbol{Y} - \widetilde{\boldsymbol{Z}}\hat{\beta}_{\mathrm{GS}}).$$

因此, 可得到 $a_j(u)$ 的估计 $\hat{a}_j(u)$, 即

$$\hat{a}_j(u) = \boldsymbol{g}_{j,k_j}^{\mathrm{T}}(u)\hat{\alpha}_{j,k_j}, \quad j = 1, \cdots, p.$$

2. **主要结果**

在给出主要结果之前, 首先给出一个定义和一些条件.

定义6.2.1 设 $\mathcal{G}$ 是一个函数类. 如果 $\mathcal{G}$ 中任一函数 $g(u, x)$ 满足:

(i) 对某个连续函数 $h_j(u)$, $g(u, x) = h^{\mathrm{T}}(u)x \equiv \sum_{j=1}^{p} x_j h_j(u)$, 其中 $h(u) = (h_1(u), \cdots, h_p(u))^{\mathrm{T}}$;

(ii) $\sum_{j=1}^{p} E[x_{ij}^2 h_j^2(u_i)] < \infty$, 其中 $x_j(x_{ij})$ 是 $x(x_j)$ 的第 j 个分量,

则称 $\mathcal{G}$ 是一个变系数函数类.

对任何函数 $f(u,x)$, 用 $E_{\mathcal{G}}[f(u,x)]$ 表示 $f(u,x)$ 到变系数函数空间 $\mathcal{G}$ 的投影 (在 L_2 模下). 也就是说 $E_{\mathcal{G}}[f(u,x)]$ 是属于 $\mathcal{G}$ 的元素, 且它的所有元素中最靠近 $f(u,x)$ 的元素. 更特别地,

$$
\begin{aligned}
&E\{(f(u,x)-E_{\mathcal{G}}[f(u,x)])(f(u,x)-E_{\mathcal{G}}[f(u,x)])^{\mathrm{T}}\}\\
&=\inf_{\sum_j x_jh_j(u)\in\mathcal{G}} E\left\{\left(f(u,x)-\sum_{j=1}^{p}x_jh_j(u)\right)\left(f(u,x)-\sum_{j=1}^{p}x_jh_j(u)\right)^{\mathrm{T}}\right\}.
\end{aligned}
$$

因此, 对所有 $g(u,x)=\sum_{j=1}^{p}x_jh_j(u)\in\mathcal{G}$,

$$
\begin{aligned}
&E\{(f(u,x)-E_{\mathcal{G}}[f(u,x)])(f(u,x)-E_{\mathcal{G}}[f(u,x)])^{\mathrm{T}}\}\\
&\leqslant E\left\{\left(f(u,x)-\sum_{j=1}^{p}x_jh_j(u)\right)\left(f(u,x)-\sum_{j=1}^{p}x_jh_j(u)\right)^{\mathrm{T}}\right\}.
\end{aligned}
$$

这里对平方矩阵 A 和 B, $A\leqslant B$ 意味着 $A-B$ 是半正定的.

下列条件将用来建立 $\hat\beta_{\mathrm{GS}}$ 的渐近正态性和 $\hat a(u)$ 的收敛速度.

条件 1 (i) 样本 $\{(U_i,X_i,Z_i,Y_i),1\leqslant i\leqslant n\}$ 是独立同分布的, 并与 (U,X,Z,Y) 有共同分布, 且 (U,X,Z) 的支撑是 $\mathbf{R}^{p+q+1}$ 中的紧子集; (ii) $E(Z|U=u,X=x)$ 和 $\mathrm{var}(Y|U=u,X=x,Z=z)$ 都是 (U,X,Z) 的支撑上的有界函数.

条件 2 (i) 对每个 K, 存在非奇异矩阵 B, 使得对 $G_K(u,x)=Bg_K(u,x)$, $E[G_K(U_i,X_i)G_K(U_i,X_i)^{\mathrm{T}}]$ 的最小特征值在 K 上一致有界且大于零; (ii) 存在满足 $\sup\limits_{(u,x)\in\mathcal{S}}|G_K(u,x)|\leqslant\zeta_0(K)$ 和 $K=K_n$ 的常数序列 $\zeta_0(K)$, 使得 $\zeta_0^2(K)K/n\to 0\ (n\to\infty)$, 其中 $\mathcal{S}$ 是 (U,X) 的支撑, 且 $|A|=[\mathrm{tr}(A^{\mathrm{T}}A)]^{1/2}$ 表示矩阵 A 的 Euclid 模.

条件 3 对 $f(u,x)=\sum_{j=1}^{p}x_jh_j(u)$, 存在 $\delta_j>0\ (j=1,\cdots,p)$ 和 $\alpha_f=\alpha_{fK}=(\alpha_{1,k_1}^{\mathrm{T}},\cdots\alpha_{p,k_p}^{\mathrm{T}})^{\mathrm{T}}$, 使得 $\sup\limits_{(u,x)\in\mathcal{S}}|f(u,x)-G_K(u,x)^{\mathrm{T}}\alpha_f|=O\left(\sum_{j=1}^{p}k_j^{-\delta_j}\right)$; (ii) 对 $\min\{k_1,\cdots,k_p)\}\to\infty$, $\sqrt{n}\sum_{j=1}^{p}k_j^{-2\delta_j}\to 0(n\to\infty)$.

条件 1 是使用在序列估计方法中的标准假定. 条件 2 通常蕴涵 (U,X) 的密度函数的下确界大于一个正的常数. 条件 3 说明存在 $\delta_j>0\ (j=1,\cdots,p)$, 使得函数一致逼近的速度为 $\sum_{j=1}^{p}k_j^{-\delta_j}$. 条件 2 和 3 不是最弱的, 但有许多序列函数满足这两

个条件, 如幂级数和样条.

在上述条件下, 可以得到下列主要定理.

定理6.2.2 记 $e = Z - E_{\mathcal{G}}(Z)$, 且假设 $\Phi \equiv E(ee^T)$ 是正定的, 则在条件 1~3 下, 有

(i) $\sqrt{n}(\hat{\beta}_{\mathrm{GS}} - \beta) \xrightarrow{D} N(0, \Sigma_0)$, 其中 $\Sigma_0 = \Phi^{-1}\Omega\Phi^{-1}$, $\Omega = E[\sigma^2(U,X,Z)ee^{\mathrm{T}}]$, $\sigma^2(U,X,Z) = E(\varepsilon^2|U,X,Z)$.

(ii) $\hat{\Sigma}_0$ 的相合估计是 $\hat{\Sigma}_0 = \hat{\Phi}^{-1}\hat{\Omega}\hat{\Phi}^{-1}$, 其中 $\hat{\Phi} = n^{-1}\sum_{i=1}^{n}(Z_i - \widetilde{Z}_i)(Z_i - \widetilde{Z}_i)^{\mathrm{T}}$, $\hat{\Omega} = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2(Z_i - \widetilde{Z}_i)(Z_i - \widetilde{Z}_i)^{\mathrm{T}}$, $\widetilde{Z}_i$ 是 $\widetilde{Z}$ 的第 i 列, $\hat{\varepsilon}_i = Y_i - \mathrm{g}_K(U_i, X_i)\hat{\alpha} - \hat{\beta}_{\mathrm{GS}}^{\mathrm{T}}Z_i$.

由 Chamberlain(1992) 的结果可知, β_0 的估计量的渐近方差的逆的半参数有效界是

$$J_0 = \inf_{g\in\mathcal{G}} E\{[Z - g(U,X)][\mathrm{var}(\varepsilon|U,X,Z)]^{-1}[Z - g(U,X)]^{\mathrm{T}}\}.$$

在误差是条件同方差下, 亦即 $\mathrm{var}(\varepsilon|U,X,Z) = \sigma^2$ 下, 上式可重新写作

$$\begin{aligned}
J_0 &= \frac{1}{\sigma^2}\inf_{g\in\mathcal{G}} E\{[Z - g(U,X)][Z - g(U,X)]^{\mathrm{T}}\} \\
&= \frac{1}{\sigma^2}\inf_{g\in\mathcal{G}} E\{[Z - E_{\mathcal{G}}(Z)][Z - E_{\mathcal{G}}(Z)]^{\mathrm{T}}\} \\
&= \frac{1}{\sigma^2}E(ee^{\mathrm{T}}) = \frac{\Phi}{\sigma^2}.
\end{aligned}$$

注意到上式的逆与 $\Sigma_0 = \sigma^2\Phi^{-1}$ 一致, 当误差是条件同方差时, J_0^{-1} 正是 $\sqrt{n}(\hat{\beta}_{\mathrm{GS}} - \beta_0)$ 的渐近方差. 因此, $\Sigma_0^{-1} = J_0$, 且在条件同方差误差假定下 $\hat{\beta}_{\mathrm{GS}}$ 是半参数有效估计.

下面的定理给出了 $\hat{a}_j(u)$ 收敛到 $a_j(u)$ 的速度.

定理6.2.3 在条件 1~3 下, 对 $j = 1, \cdots, p$, 有

(i) $\sup_{u\in\mathcal{U}}|\hat{a}_j(u) - a_j(u)| = O_P\left(\zeta_0(K)\left(\sqrt{K}/\sqrt{n} + \sum_{j=1}^{p}k_j^{-\delta_j}\right)\right)$, 其中 $\mathcal{U}$ 表示 U 的支撑.

(ii) $\frac{1}{n}\sum_{j=1}^{n}[\hat{a}_j(u) - a_j(u)]^2 = O_P\left(K/n + \sum_{j=1}^{p}k_j^{-2\delta_j}\right)$.

(iii) $\int[\hat{a}_j(u) - a_j(u)]^2\mathrm{d}F_Z(z) = O_P\left(K/n + \sum_{j=1}^{p}k_j^{-2\delta_j}\right)$, 其中 $F_Z(z)$ 是 Z 的分布函数.

定理 6.2.2 和定理 6.2.3 的证明可参见文献 (Ahmad et al., 2005).

6.3 profile 似然比检验

本节主要阐述部分线性变系数模型的拟合优度检验问题. 对模型中参数分量构造 profile 似然比检验和 Wald 检验统计量, 对非参数分量构造广义似然比检验统计量, 并研究其渐近性质. 本节的主要内容取材于 (Fan and Huang, 2005).

6.3.1 profile 似然比统计量

模型中参数分量和非参数分量的估计可以用来拟合模型, 这是统计推断的任务之一. 统计推断的另一个任务是模型检验. 也就是说, 拟合模型以后, 一个自然的任务是检验参数部分一个确定的变量是否统计显著. 这就引入检验问题

$$H_0 : \beta_1 = \cdots = \beta_l = 0, \quad l \leqslant q.$$

更一般地, 人们可以考虑线性假设

$$H_0 : A\beta = 0, \tag{6.3.1}$$

其中 A 是一个 $l \times q$ 满秩矩阵. 这是一个半参数假设对另一个半参数假设检验问题. 为构造 profile 似然比统计量, 暂且假定 $\varepsilon \sim N(0, \sigma^2)$. 在模型 6.2.1 下, 似然函数为

$$l(\mathbf{a}, \beta, \sigma) = -n\log(\sqrt{n}\sigma) - \mathrm{RSS}_1^*/(2\sigma^2),$$

其中 $\mathrm{RSS}_1^* = \sum\limits_{i=1}^{n}[Y_i - \beta^{\mathrm{T}} Z_i - \boldsymbol{a}^{\mathrm{T}}(U_i)X_i]^2$. 对给定的 β, 利用局部线性拟合可构造 $\mathbf{a}(\cdot)$ 的估计量 $\hat{\boldsymbol{a}}(\cdot;\beta)$. 将它代入上述似然函数中可得

$$l(\hat{\boldsymbol{a}}(\cdot;\beta), \beta, \sigma) = -n\log(\sqrt{n}\sigma) - \mathrm{RSS}_1^{**}/(2\sigma^2), \tag{6.3.2}$$

其中 $\mathrm{RSS}_1^{**} = \sum\limits_{i=1}^{n}[Y_i - \beta^{\mathrm{T}} Z_i - \hat{\boldsymbol{a}}^{\mathrm{T}}(U_i;\beta)X_i]^2$. 对 (6.3.2) 关于 β 和 σ^2 极大化可得到如 (6.2.5) 式给出的 β 的 profile 似然估计量, 仍记为 $\hat{\beta}_{\mathrm{P}}$, 以及 σ^2 的估计量 $\hat{\sigma}^2 = n^{-1}\mathrm{RSS}_1$, 其中

$$\mathrm{RSS}_1 = \sum_{i=1}^{n}[Y_i - \hat{\beta}_{\mathrm{P}}^{\mathrm{T}} Z_i - \hat{\boldsymbol{a}}^{\mathrm{T}}(U_i;\hat{\beta}_{\mathrm{P}})X_i]^2. \tag{6.3.3}$$

注意到 $\hat{\beta}_{\mathrm{P}}$ 和 $\hat{\sigma}^2$ 不是极大似然估计, 因为 $\hat{\boldsymbol{a}}(\cdot;\beta)$ 不是由极大似然方法得到. 将这些估计量代入 (6.3.2) 式可得到 profile 似然

$$l_n(H_1) = -(n/2)\log(2\pi/n) - (n/2)\log(\mathrm{RSS}_1) - n/2.$$

另一方面, 在零假设 (6.3.1) 下, profile 似然估计量是在限制 (6.3.1) 下极大化 (6.3.2) 的解. 此解与 "合成线性模型" 的限制的最小二乘问题相同, 那就是

$$\hat{\beta}_0 = \hat{\beta}_{\mathrm{P}} - (\widetilde{\boldsymbol{Z}}^{\mathrm{T}}\widetilde{\boldsymbol{Z}})^{-1}A^{\mathrm{T}}[A^{\mathrm{T}}(\widetilde{\boldsymbol{Z}}^{\mathrm{T}}\widetilde{\boldsymbol{Z}})^{-1}A]^{-1}A\hat{\beta}_{\mathrm{P}},$$

其中 $\widetilde{\boldsymbol{Z}} = (I - S)\boldsymbol{Z}$. 因此, 在零假设下对 (6.3.2) 关于 β 和 σ^2 极大化也可得到 profile 似然

$$l_n(H_0) = -(n/2)\log(2\pi/n) - (n/2)\log(\mathrm{RSS}_0) - n/2,$$

其中

$$\mathrm{RSS}_0 = \sum_{i=1}^{n}[Y_i - \hat{\beta}_0^{\mathrm{T}}Z_i - \hat{\boldsymbol{a}}^{\mathrm{T}}(U_i;\hat{\beta}_0)X_i]^2.$$

现在, 利用上面得到的 profile 似然可以定义广义似然统比计量为

$$T_n = l_n(H_1) - l_n(H_0) = \frac{n}{2}\log\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1} \approx \frac{n}{2}\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}. \tag{6.3.4}$$

这个广义似然比统计量称作 profile 似然比统计量. 正如前面提及的, 这与极大似然比检验不同, 因为 profile 似然估计量没有极大化似然函数. 注意到在零假设和对立假设下使用同样的带宽构造 profile 似然估计量. 这就是此方法的成功所在.

6.3.2 Wilks 现象

相对于极大似然比检验可以近似地得到 profile 似然比检验统计量. 然而, 它们也是非常困难的. 讨厌函数全是非参数, 因此是无限维的. 不能使用极大似然方法估计. 这个问题提升到是否渐近零分布仍服从 χ^2 分布. 下面的定理回答了这个问题: 传统的似然理论仍适用.

定理6.3.1 设条件 (C6.2.1)~(C6.2.5) 成立, 则在零假设下, $2T_n \xrightarrow{D} \chi_l^2$.

定理 6.3.1 针对广义似然比检验揭露了一类新的 Wilks 现象, 它为检验半参数模型中的参数分量提供了一个有用的结果. 这样一个有吸引力的 Wilks 类现象允许人们使用渐近分布来计算临界值, 其中讨厌参数的估计将从零模型中得到, 即对模型 (6.2.1) 不需要假定 (6.3.1) 式, 无论零分布对特殊数据是否为真都是如此.

6.3.3 profile 似然比检验的功效

下面提供一个计算 profile 似然比检验的功效的计算公式. 假设

$$n\beta_n^{\mathrm{T}}A^{\mathrm{T}}(A\Sigma A)^{-1}A\beta_n \to \lambda,$$

其中 $\sqrt{n}\beta_n \to 0$. 上式不仅允许我们确定半参数检验问题的样本量, 而且可以与下面将要介绍的 Wilks 类检验功效作比较.

定理6.3.2 设条件 (C6.2.1)~(C6.2.5) 成立, 则在检验问题 (6.3.1) 的对立假设下, $2T_n \xrightarrow{D} \chi_l^2(\lambda)$, 其中 $\chi_l^2(\lambda)$ 表示自由度为 l 的非中心 χ^2 分布, $\lambda = n\beta^{\mathrm{T}}A^{\mathrm{T}}(A\Sigma A)^{-1}A\beta$, Σ 在定理 6.2.1 中定义.

对给定的 β 的值, 功效的计算需要估计 Σ. 它的估计量将在下面的 (6.3.6) 或 (6.3.7) 式中给出.

6.3.4 Wald 检验

检验问题 (6.3.1) 也可以使用下面的 Wald 统计量:

$$W_n = \hat{\beta}_{\mathrm{P}}^{\mathrm{T}} A^{\mathrm{T}} (A\hat{\Sigma}A)^{-1} A\hat{\beta}_{\mathrm{P}}, \tag{6.3.5}$$

其中 $\hat{\beta}_{\mathrm{P}}$ 在 (6.2.5) 式中定义, $\hat{\Sigma}$ 是 $\hat{\beta}_{\mathrm{P}}$ 的协方差阵的估计. 遵照 (6.2.5) 式, $\hat{\Sigma}$ 可定义为

$$\hat{\Sigma} = \hat{\sigma}^2(\widetilde{\boldsymbol{Z}}^{\mathrm{T}}\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}^{\mathrm{T}}(I-S)(I-S)^{\mathrm{T}}\widetilde{\boldsymbol{Z}}(\widetilde{\boldsymbol{Z}}^{\mathrm{T}}\widetilde{\boldsymbol{Z}})^{-1}, \tag{6.3.6}$$

其中 $\hat{\sigma}^2$ 是残差的样本方差. 一个简单的协方差阵的估计为

$$\hat{\Sigma}^* = \hat{\sigma}^2(\widetilde{\boldsymbol{Z}}^{\mathrm{T}}\widetilde{\boldsymbol{Z}})^{-1}. \tag{6.3.7}$$

可以证明 $\hat{\Sigma}$ 和 $\hat{\Sigma}^*$ 都是 Σ 的相合估计. profile 似然比统计量的优点是不包含估计的协方差阵, 这样就不需要首先研究 $\hat{\beta}_{\mathrm{P}}$ 的渐近正态性.

基于半参数有效估计 $\hat{\beta}_{\mathrm{P}}$ 的 Wald 类检验统计量 (6.3.5) 是有魅力的, 且可作为其他方法的基准. 下列定理给出了渐近零分布和功效.

定理6.3.3 设条件 (C6.2.1)~(C6.2.5) 成立, 则在零假设下, $W_n \xrightarrow{D} \chi_l^2$. 在检验问题 (6.3.1) 的对立假设下, $W_n \xrightarrow{D} \chi_l^2(\lambda)$, 其中 $\lambda = n\beta^{\mathrm{T}}A^{\mathrm{T}}(A\Sigma A)^{-1}A\beta$, Σ 在定理 6.2.1 中定义.

定理 6.3.3 表明 Wald 类统计量具有与 profile 似然比统计量同样的渐近零分布和对立分布. 这就给出了 profile 似然比统计量的理论依据. 另一方面, profile 似然比统计量对许多实际问题的使用是最简单的.

6.3.5 非参数分量的检验

得到 $\{a_1(\cdot), \cdots, a_p(\cdot)\}$ 的非参数估计以后, 常见的研究要求非参数分量是否可以用参数模型拟合. 这就诱使我们考虑假设检验问题

$$H_0: a_p(u) = a_p(u;\theta) \quad 对 \quad H_0: a_p(u) \neq a_p(u;\theta),$$

其中 θ 是一个未知的参数向量. 上述问题包括检验变量 X_p 的显著性, 视 $a_p(u;\theta) = 0$ 或 $a_p(u;\theta) = a_p$ 而定. 可以类似地处理其他检验问题.

为表达简单, 考虑检验问题

$$a_1(u;\theta) = a_1, \cdots, a_p(u;\theta) = a_p. \tag{6.3.8}$$

这个思想可以容易地推广到更一般情况.

设 $\tilde{\boldsymbol{a}} = (\tilde{a}_1, \cdots, \tilde{a}_p)^{\mathrm{T}}$ 和 $\tilde{\beta}$ 是 H_0 下的最小二乘估计. 遵照 Fan 和 Li(2001), 一个广义似然比统计量定义为

$$T_0 = \frac{n}{2}\log\frac{\widetilde{\mathrm{RSS}}_0}{\mathrm{RSS}_1},$$

其中 RSS_1 在 (6.3.3) 中定义,

$$\widetilde{\mathrm{RSS}}_0 = \sum_{i=1}^{n}[Y_i - \tilde{\beta}^{\mathrm{T}}Z_i - \tilde{\boldsymbol{a}}^{\mathrm{T}}X_i]^2.$$

下列定理给出了 T_0 的渐近分布.

定理6.3.4 设条件 (C6.2.1)~(C6.2.5) 成立. 如果 $h \to 0$, $nh^{3/2} \to \infty$, 则在零假设下, $r_K T_0 \xrightarrow{D} \chi^2_{\delta_n}$, 其中

$$r_K = \frac{K(0) - 0.5\displaystyle\int K^2(t)\mathrm{d}t}{\displaystyle\int [K(t) - 0.5K * K(t)]^2\mathrm{d}t},$$

$$\delta_n = r_K\frac{p|\mathcal{U}|}{h}\left[K(0) - 0.5\int K^2(t)\mathrm{d}t\right],$$

$|\mathcal{U}|$ 是 U 的支撑的长度.

6.3.6 带宽选择

profile 似然估计量依赖于带宽的选择. 进一步, profile 似然比统计量和 Wald 统计量也包括带宽选择. 在实际中自然出现带宽选择的问题.

正如 Fan 和 Li(2001) 讨论的那样, 由于在非参数分量的估计中带宽的选择没有产生敏感的偏差, 因此参数分量的估计不是很灵敏地依赖于带宽的选择. 理由是在参数分量的估计过程中非参数分量估计中的偏差没有达到平均数, 非参数估计中的方差达到了平均数.

基于上述原因, 我们提出使用交叉验证方法确定光滑参数的近似值. 假设 $\hat{a}_h^{(-i)}(\cdot)$ 和 $\hat{\beta}_h^{(-i)}$ 是去掉第 i 个观察的 profile 似然估计, 其计算可以参看 (6.2.4) 和 (6.2.5). 定义交叉验证函数

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{\beta}_h^{(-i)\mathrm{T}}Z_i - \hat{a}_h^{(-i)\mathrm{T}}(U_i)X_i\right]^2. \tag{6.3.9}$$

依赖于光滑参数 h 的 (6.3.9) 式用作估计方案的有效性的整体测量. 最小二乘交叉验证带宽定义为 $\hat{h}_{\mathrm{CV}} = \arg\min_h \mathrm{CV}(h)$.

6.4 变量选择

6.4.1 引言

对模型 (6.1.1) 的变量选择问题, Li 和 Liang(2008) 利用 SCAD 变量选择方法研究了参数分量的变量选择问题. 其基本思想是把模型 (6.1.1) 转化为如下线性模型 $Y^* = \beta_0^{\mathrm{T}} Z + \varepsilon$, 其中 $Y^* = Y - \tilde{a}^{\mathrm{T}}(U)X$, $\tilde{a}(\cdot)$ 为 $a(\cdot)$ 的相合估计. 从而, 利用 SCAD 变量选择方法来选择参数分量中的重要变量, 并给出 β_0 的正则估计 $\hat{\beta}$. 然后, 用 $\hat{\beta}$ 代替模型 (6.1.1) 中的 β_0, 并基于广义似然比检验 (GLRT) 来选择非参数分量中的重要变量. 尽管该变量选择过程在理论上是可行的, 但是在实际应用中有以下几点需要考虑. 第一, 利用 GLRT 方法对非参数分量进行变量选择往往是相当困难的. 因为, 对每一个子模型都需要给出其中函数系数的非参数估计, 这将大大增加计算量. 另外, GLRT 的渐近分布为带有扩散自由度的 χ^2 分布, 因此, 获得 GLRT 检验的临界值往往也是相当困难的. 第二, 在对参数分量的变量选择过程中, 当 $a(\cdot)$ 的维数非常高时, 除非样本容量 n 充分大, 否则很难给出 $a(\cdot)$ 的有效估计.

Zhao 和 Xue(2009a) 利用基函数逼近和惩罚最小二乘方法, 对模型 (6.1.1) 提出了一种变量选择方法. 通过选择合适的调整参数, 证明了所提出的变量选择方法可以相合地识别出真实模型, 且所得回归系数的正则估计具有 oracle 性质. 这里 oracle 性质表示对非参数分量的估计达到了最优的收敛速度, 且对非零参数分量的估计与利用真实模型所给的估计具有相同的渐近分布. Zhao 和 Xue(2009a) 的方法具有两个优点: 一是他们的方法对非参数分量的变量选择和估计可以同时完成, 这就减少了计算量; 二是他们的方法可以同时对参数分量和非参数分量进行变量选择. 本节的内容取材于文献 (Zhao and Xue, 2009a).

6.4.2 方法论和主要结果

设 $B(u) = (B_1(u), \cdots, B_L(u))^{\mathrm{T}}$ 为 $M+1$ 阶 B 样条基函数, 其中 $L = K+M+1$, K 为内部节点个数. 那么, $a_j(u)(j = 1, \cdots, p)$ 可以由下式逼近:

$$a_j(u) \approx \gamma_j^{\mathrm{T}} B(u), \quad j = 1, \cdots, p. \tag{6.4.1}$$

因此, 定义惩罚最小二乘目标函数为

$$Q(\gamma, \beta) = \sum_{i=1}^{n} \left[Y_i - \beta^{\mathrm{T}} Z_i - \gamma^{\mathrm{T}} W_i\right]^2 + n \sum_{j=1}^{p} p_{\lambda_{1j}}(\|\gamma_j\|_H) + n \sum_{l=1}^{q} p_{\lambda_{2l}}(|\beta_l|), \tag{6.4.2}$$

其中 $W_i = I_p \otimes B(U_i) \cdot X_i$, $\gamma = (\gamma_1^{\mathrm{T}}, \cdots, \gamma_p^{\mathrm{T}})^{\mathrm{T}}$, $\|\gamma_j\|_H = (\gamma_j^{\mathrm{T}} H \gamma_j)^{1/2}$, $H = \int B(u) B^{\mathrm{T}}$

$(u)\mathrm{d}u$, $p_\lambda(\cdot)$ 为惩罚函数, λ 为调整参数. 此处采用 SCAD 惩罚函数:

$$p'_\lambda(w) = \lambda\left\{I(w \leqslant \lambda) + \frac{(a\lambda - w)_+}{(a-1)\lambda} I(w > \lambda)\right\},$$

其中 $a > 2$, $w > 0$, $p_\lambda(0) = 0$. 这里, 对应 $a_j(\cdot)$ 和 β_l 调整参数 λ_{1j} 和 λ_{2l} 没必要完全相同. 这样我们就可以对非零回归系数选择较小的调整参数, 而对零回归系数选择较大的调整参数. 因此, 可以保证在得到非零回归系数的相合估计的同时可以把零回归系数的估计压缩为 0, 进而达到变量选择的目的.

记 $\hat\beta$ 和 $\hat\gamma = (\hat\gamma_1^{\mathrm{T}}, \cdots, \hat\gamma_p^{\mathrm{T}})^{\mathrm{T}}$ 为极小化 (6.4.2) 式所得的解. 那么, $\hat\beta$ 为 β_0 的惩罚最小二乘估计, 而 $a_j(u)$ 的估计量为 $\hat a_j(u) = \hat\gamma_j^{\mathrm{T}} B(u)$, 且 $\hat{\boldsymbol a}(u) = (\hat a_1(u), \cdots, \hat a_p(u))^{\mathrm{T}}$.

下面研究估计量 $\hat\beta$ 和 $\hat{\boldsymbol a}(u)$ 的渐近性质. 为此, 首先给出一些记号. 令 $\boldsymbol a_0(\cdot)$ 为 $\boldsymbol a(\cdot)$ 的真值. 不失一般性, 假定 $\{a_{j0}(\cdot) = 0,\ j = d+1, \cdots, p\}$, 且 $\{a_{j0}(\cdot),\ j = 1, \cdots, d\}$ 为 $\boldsymbol a_0(\cdot)$ 的所有非零分量. 进一步, 假定 $\{\beta_{l0} = 0,\ l = s+1, \cdots, q\}$, 且 $\{\beta_{l0},\ l = 1, \cdots, s\}$ 为 β_0 的所有非零分量.

在给出本节的主要定理之前, 首先给出一组条件.

(C6.4.1) $\sup\limits_{0\leqslant u\leqslant 1} (\varepsilon^4|U = u) < \infty$, $\sup\limits_{0\leqslant u\leqslant 1} (||X||^4|U = u) < \infty$.

(C6.4.2) 对任给的点 $u \in (0,1)$, 函数 $\mathbf{a}(u)$ 和 $f(u)$ 皆在点 u 处 $r(> 1/2)$ 次连续可微, $\sigma^2(u)$ 为 u 的连续函数, 其中 $f(u)$ 为 U 的密度函数, $\sigma^2(u) = E\{\varepsilon^2|U = u\}$.

(C6.4.3) 令 $c_1, \cdots, c_K$ 为区间 $[0,1]$ 上的内部节点, 且令 $c_0 = 0$, $c_{K+1} = 1$ 和 $h_i = c_i - c_{i-1}$, 那么, 存在某一常数 C_0 满足

$$\frac{\max\{h_i\}}{\min\{h_i\}} \leqslant C_0, \quad \max\{|h_{i+1} - h_i|\} = o\left(K^{-1}\right).$$

(C6.4.4) $b_n \to 0$, 其中 $b_n = \max\limits_{j,l}\left\{|p''_{\lambda_{2l}}(|\beta_{l0}|)|,\ |p''_{\lambda_{1j}}(||\gamma_{j0}||_H)| : \beta_{l0} \neq 0,\ \gamma_{j0} \neq 0\right\}$.

(C6.4.5) $\liminf\limits_{n\to\infty} \liminf\limits_{\beta_l\to 0^+} \lambda_{2l}^{-1} p'_{\lambda_{2l}}(|\beta_l|) > 0$, $\liminf\limits_{n\to\infty} \liminf\limits_{||\gamma_j||_H\to 0} \lambda_{1j}^{-1} p'_{\lambda_{1j}}(||\gamma_j||_H) > 0$, $l = s+1, \cdots, q$, $j = d+1, \cdots, p$.

上述条件在非参数估计和变量选择中是常见的正则条件. (C6.4.1) 和 (C6.4.2) 是常用的矩条件和函数光滑条件. 条件 (C6.4.3) 表明 $c_0, \cdots, c_{j+1}$ 为区间 $[0,1]$ 的一个 C_0 拟均匀分隔序列. 条件 (C6.4.4) 和 (C6.4.5) 是关于惩罚函数的假设条件, 这些条件类似于 Fan 和 Li(2001) 与 Li 和 Liang(2008) 中所用的条件.

下列定理给出了 $\hat\beta$ 和 $(\hat a_1(\cdot), \cdots, \hat a_p(\cdot))^{\mathrm{T}}$ 的相合性.

定理6.4.1 设条件(C6.4.1)~(C6.4.5)成立, 并且内部节点个数$K = O(n^{1/(2r+1)})$, 则

$$||\hat\beta - \beta_0|| = O_p(n^{\frac{-r}{2r+1}} + c_n),$$

$$||\hat{a}_j(\cdot) - a_{j0}(\cdot)|| = O_p(n^{\frac{-r}{2r+1}} + c_n), \quad j = 1, \cdots, p,$$

其中 $c_n = \max_{j,l}\left\{|p'_{\lambda_{2l}}(|\beta_{l0}|)|,\ |p'_{\lambda_{1j}}(||\gamma_{j0}||_H)| : \beta_{l0} \neq 0,\ \gamma_{j0} \neq 0\right\}$, r 在条件 (C6.2.2) 中定义.

下列定理表明, 在一些条件下, 本节所给出的正则估计具有稀疏性.

定理6.4.2 设条件 (C6.4.1)~(C6.4.5) 成立, 且内部节点个数 $K = O(n^{1/(2r+1)})$. 记 $\lambda_{\max} = \max_{j,l}\{\lambda_{1j}, \lambda_{2l}\}$, $\lambda_{\min} = \min_{j,l}\{\lambda_{1j}, \lambda_{2l}\}$. 如果 $\lambda_{\max} \to 0$, $n^{r/(2r+1)}\lambda_{\min} \to \infty$, 则当 $n \to \infty$ 时, $\hat{\beta}$ 和 $\hat{a}(\cdot)$ 以概率 1 满足

(i) $\hat{\beta}_l = 0, l = s+1, \cdots, q$,

(ii) $\hat{a}_j(\cdot) \equiv 0,\ j = d+1, \cdots, p$.

由 (Fan and Li, 2001) 的注 1 可知, 当 $n \to \infty$ 时, 如果 $\lambda_{\max} \to 0$, 那么 $a_n = 0$. 因此由定理 6.4.1 和定理 6.4.2 可知, 通过选择合适的调整参数, 提出的变量选择方法可以相合地识别出真实模型, 且非参数分量的估计达到了最优的收敛速度 (Stone, 1982). 接下来我们证明对 β_0 中非零分量的估计与在已知真实模型的情况下所给的估计具有相同的渐近分布. 为此, 需要另外的记号. 令 $\beta^* = (\beta_1, \cdots, \beta_s)^{\mathrm{T}}$, $a^*(u) = (a_1(u), \cdots, a_d(u))^{\mathrm{T}}$, 并且用 β_0^* 和 $a_0^*(u)$ 分别表示 β^* 和 $a^*(u)$ 的真值. 对应的协变量分别记为 Z^* 和 X^*. 另外, 记

$$\Sigma^* = E\left(Z^* Z^{*\mathrm{T}}\right) - E\{E\left(Z^* X^{*\mathrm{T}}|U\right)\left[E\left(X^* X^{*\mathrm{T}}|U\right)\right]^{-1} E\left(X^* Z^{*\mathrm{T}}|U\right)\}.$$

假定 Σ 为可逆矩阵. 下列定理给出了 $\hat{\beta}^*$ 的渐近正态性.

定理6.4.3 设条件 (C6.4.1)~(C6.4.5) 成立, 且内部节点个数 $K = O(n^{1/(2r+1)})$, 则

$$\sqrt{n}(\hat{\beta}^* - \beta_0^*) \xrightarrow{D} N(0, \sigma^2 \Sigma^{*-1}),$$

其中 $\sigma^2 = E(\varepsilon^2)$.

定理 6.4.1 和定理 6.4.3 表明所得的惩罚最小二乘估计具有 oracle 性质. 也就是说, 对非参数分量的估计达到了最优的收敛速度, 并且非零参数分量估计的渐近分布与基于真实模型所给估计的渐近分布是相同的.

6.4.3 计算方法

因为 $Q(\gamma, \beta)$ 中的惩罚函数在原点是奇异的, 所以普通的梯度计算方法将不能直接应用. 类似 Fan 和 Li(2001), 利用对惩罚函数 $p_\lambda(\cdot)$ 进行二次逼近的方法来给出一个迭代计算程序. 具体地, 在任一给定的非零 w_0 的某个小邻域内, 惩罚函数 $p_\lambda(\cdot)$ 在 w_0 点可以近似表示为

$$p_\lambda(|w|) \approx p_\lambda(|w_0|) + \frac{p'_\lambda(|w_0|)}{2|w_0|}(w^2 - w_0^2).$$

因此, 对给定的初始值 $\beta_l^{(0)}$ 和 $\gamma_j^{(0)}$, 且满足 $|\beta_l^{(0)}|>0$, $l=1,\cdots,q$, $\|\gamma_j^{(0)}\|_H>0$, $j=1,\cdots,p$, 可以得到

$$p_{\lambda_{2l}}(|\beta_l|)\approx p_{\lambda_{2l}}(|\beta_l^{(0)}|)+\frac{p'_{\lambda_{2l}}(|\beta_l^{(0)}|)}{2|\beta_l^{(0)}|}\left(|\beta_l|^2-|\beta_l^{(0)}|^2\right), \tag{6.4.3}$$

$$p_{\lambda_{1j}}(\|\gamma_j\|_H)\approx p_{\lambda_{1j}}(\|\gamma_j^{(0)}\|_H)+\frac{p'_{\lambda_{1j}}(\|\gamma_j^{(0)}\|_H)}{2\|\gamma_j^{(0)}\|_H}\left(\|\gamma_j\|_H^2-\|\gamma_j^{(0)}\|_H^2\right). \tag{6.4.4}$$

记 $\widetilde{Z}_i=(Z_i^{\mathrm{T}},W_i^{\mathrm{T}})^{\mathrm{T}}$, $\alpha=(\beta^{\mathrm{T}},\gamma^{\mathrm{T}})^{\mathrm{T}}$,

$$\Sigma(\alpha^{(0)})=\operatorname{diag}\left\{\frac{p'_{\lambda_{21}}(|\beta_1^{(0)}|)}{|\beta_1^{(0)}|},\cdots,\frac{p'_{\lambda_{2q}}(|\beta_q^{(0)}|)}{|\beta_q^{(0)}|},\frac{p'_{\lambda_{11}}(\|\gamma_1^{(0)}\|_H)}{\|\gamma_1^{(0)}\|_H}H,\cdots,\frac{p'_{\lambda_{1p}}(\|\gamma_p^{(0)}\|_H)}{\|\gamma_p^{(0)}\|_H}H\right\}.$$

结合 (6.4.3) 和 (6.4.4) 式可知, 除了一个常数项, (6.4.2) 式可写为

$$Q(\alpha)=\sum_{i=1}^n\{Y_i-\widetilde{Z}_i^{\mathrm{T}}\alpha\}^2+\frac{n}{2}\alpha^{\mathrm{T}}\Sigma(\alpha^{(0)})\alpha. \tag{6.4.5}$$

注意到 (6.4.5) 式为一个二次型, 并且可以通过下式求解:

$$\left(\sum_{i=1}^n\widetilde{Z}_i\widetilde{Z}_i^{\mathrm{T}}+\frac{n}{2}\Sigma(\alpha^{(0)})\right)\alpha=\sum_{i=1}^n\widetilde{Z}_iY_i. \tag{6.4.6}$$

从而, 可以得到如下迭代计算步骤:

步骤 1. 给初始值 $\alpha^{(0)}$.

步骤 2. 令 $\alpha^{(0)}=\alpha^{(k)}$, 通过方程 (6.4.6) 解得 $\alpha^{(k+1)}$.

步骤 3. 重复步骤 2, 直到收敛, 并且记最终 α 的估计为 $\hat{\alpha}$. 那么 $\hat{\beta}=(I_{q\times q},0_{q\times pL})\hat{\alpha}$, 且 $\hat{\gamma}=(0_{pL\times q},I_{pL\times pL})\hat{\alpha}$.

在步骤 1 中, 我们可以基于 (6.4.2) 式右边的第一项, 并利用普通的最小二乘方法给出 α 的初始解. 为了利用该算法, 内部节点个数 K 以及惩罚函数中的调整参数 a, λ_{1j} 和 λ_{2l} 需要选择. Fan 和 Li(2001) 表明模型的变量选择结果对 a 的选取不是太敏感, 并建议取 $a=3.7$. 因此在本节中, 我们仍采用 $a=3.7$. 另外, 类似 Wang 等 (2009), 我们用 "交叉验证法" 选取其他调整参数. 具体地, 我们可以通过极小化下列交叉验证目标函数来估计 λ_{1j}, λ_{2l} 和 K.

$$\mathrm{CV}(K,\lambda_{11},\cdots,\lambda_{1p},\lambda_{21},\cdots,\lambda_{2q})=\sum_{i=1}^n\left\{Y_i-\widetilde{Z}_i^{\mathrm{T}}\hat{\alpha}^{(-i)}\right\}^2, \tag{6.4.7}$$

其中 $\hat{\alpha}^{(-i)}$ 为去掉第 i 个观察后, (6.4.2) 式的解. 但是, 在 $p+q+1$ 维空间上最小化 (6.4.7) 式常常是非常困难的. 因此, 在实际应用中, 我们建议取

$$\lambda_{1j}=\frac{\lambda}{\|\hat{\gamma}_j^{(0)}\|_H},\quad \lambda_{2l}=\frac{\lambda}{|\hat{\beta}_l^{(0)}|}, \tag{6.4.8}$$

其中 $\hat{\gamma}_j^{(0)}$ 和 $\hat{\beta}_l^{(0)}$ 分别为由 (6.4.2) 式右边的第一项并利用普通最小二乘法得到的 γ_j 和 β_l 的初始估计. 于是, (6.4.8) 式可转化为下列二维的极小化问题:

$$\mathrm{CV}(K,\lambda)=\sum_{i=1}^{n}\left\{Y_i-\widetilde{Z}_i^{\mathrm{T}}\hat{\alpha}^{(-i)}\right\}^2. \tag{6.4.9}$$

通过 (6.4.9) 式所给出的 λ_{1j} 和 λ_{2l} 满足对零系数的调整参数大于对非零系数的调整参数. 于是我们可以在给出非零系数相合估计的同时, 把零系数的估计压缩为 0, 从而达到变量选择的目的. 从下一节的模拟结果可以看出, 本节所提出的调整参数的选择方法是可行的.

6.4.4 模拟研究

这一节通过数值模拟来研究 6.4.3 小节所提出的变量选择方法的有限样本性质. 类似于 Li 和 Liang(2008), 我们利用广义均方误差 (GMSE) 来评价参数分量 $\hat{\beta}$ 的估计精度, 其定义为

$$\mathrm{GMSE}=(\hat{\beta}-\beta_0)^{\mathrm{T}}E(ZZ^{\mathrm{T}})(\hat{\beta}-\beta_0).$$

并且利用平均平方误差的平方根 (RASE) 来评价非参数分量 $\hat{a}(\cdot)$ 的估计精度, 其定义为

$$\mathrm{RASE}=\left\{\frac{1}{N}\sum_{s=1}^{N}\sum_{j=1}^{p}[\hat{a}_j(u_s)-a_j(u_s)]^2\right\}^{1/2},$$

其中 u_s, $s=1,\cdots,N$ 为用于计算 $\hat{a}(u)$ 的格子点. 在我们的模拟中取 $N=200$.

我们从模型 (6.2.1) 中产生数据, 其中 $p=q=10$, β_0 的前三个分量分别为 2.5, 1.2, 0.5, 其余分量都为 0; $a(u)$ 的前两个分量函数分别为 $a_1(u)=5.5+0.1\exp(2u-1)$ 和 $a_2(u)=2-u(2-u)$, 其他分量函数都为 0. 为实施模拟, 取协变量 $Z_l\sim N(0,1.5)$, $l=1,\cdots,10$, $X_j\sim N(0,1)$, $j=1,\cdots,10$, $U\sim U(0,1)$. Y 由模型产生, 其中 $\varepsilon\sim N(0,0.5)$. 在如下的模拟过程中, 我们取立方 B 样条基函数, 并且样本容量分别取 $n=100,150,200$. 另外, 在每种情况下, 实验重复 1000 次.

我们首先把本节提出的基于自适应调整参数 (ATP) 的变量选择方法与基于固定常数调整参数 (CTP) 的变量选择方法进行比较, 其中后者为对所有的回归系数用同一个固定调整参数的变量选择过程. 基于 1000 次重复实验, 对参数分量以及非参数分量的模拟结果见表 6.4.1, 其中 “C” 表示把真实零系数估计成 0 的平均个数, “I” 表示把真实非零系数估计成 0 的平均个数. 另外, 表 6.4.1 还给出了参数分量的 GMSE 的中位数和非参数分量的 RASE 的中位数.

表 6.4.1 基于自适应调整参数 (ATP) 和常数调整参数 (CTP) 的变量选择结果

		$n=100$			$n=150$			$n=200$		
β_0		C	I	GMSE	C	I	GMSE	C	I	GMSE
	ATP	6.661	0.003	0.015	6.775	0	0.012	6.874	0	0.008
	CTP	5.756	0.001	0.050	6.555	0	0.016	6.821	0	0.009
$\boldsymbol{a}(\cdot)$		C	I	RASE	C	I	RASE	C	I	RASE
	ATP	7.650	0	0.073	7.893	0	0.016	7.977	0	0.011
	CTP	7.091	0	0.101	7.584	0	0.027	7.971	0	0.012

由表 6.4.1 可知, 随着 n 的增大, 基于 ATP 和 CTP 的变量选择过程所得的最终模型越来越接近真实模型, 并且对应参数分量的 GMSE 和对应非参数分量的 RASE 均随着 n 的增大而变小. 但是, 对给定的 n, 基于 ATP 的变量选择过程优于基于 CTP 的变量选择过程, 因为后者不能消除某些不重要的变量, 并且给出相对较大的 GMSE 和 RASE. 特别是当 n 较小时, 该结论更加明显, 而对较大的 n, 这两种方法的模拟结果类似.

接下来, 我们研究内部节点个数 K 的选取对参数分量变量选择的影响. 为此, 在模拟过程中, 内部节点个数分别取为 $K=K_0$, $\lfloor 1.5K_0\rfloor$ 和 $\lfloor K_0/1.5\rfloor$, 其中 K_0 为通过 (6.4.9) 式所得的节点个数. 模拟结果见表 6.4.2, 其中 “Oracle” 表示在零回归系数已知的情况下, 基于真实模型所给出的估计, 从表 6.4.2 可以看出, 随着 n 的增加, 利用 6.4.3 小节所提出的变量选择方法对参数分量的变量选择结果越来越接近于基于 oracle 方法所得的结果. 同时还可以看出, 对给定的样本容量 n, 利用 6.4.3 小节的方法对参数分量的变量选择结果对内部节点个数的选取不是太敏感.

表 6.4.2 对不同的内部节点个数, 基于 sSCAD 方法, 对参数分量的变量选择结果, 其中 $K_1=\lfloor 1.5K_0\rfloor$, $K_2=\lfloor K_0/1.5\rfloor$

	$n=100$			$n=150$			$n=200$		
K	C	I	GMSE	C	I	GMSE	C	I	GMSE
K_0	6.665	0.003	0.016	6.776	0	0.012	6.877	0	0.008
K_1	6.621	0.001	0.019	6.761	0	0.014	6.873	0	0.009
K_2	6.627	0.034	0.024	6.754	0	0.012	6.869	0	0.009
Oracle	7	0	0.011	7	0	0.007	7	0	0.005

最后, 我们把 6.4.3 小节提出的半参数变量选择方法 (sSCAD) 与已有的变量选择方法进行比较. 对参数分量, 我们把 sSCAD 方法与 Li 和 Liang(2008) 所应用的变量选择方法 (pSCAD) 进行比较, 模拟结果见表 6.4.3. 对非参数分量, 我们把 sSCAD 方法与 Li 和 Liang(2008) 所应用广义似然比检验方法 (GLRT) 进行比较, 模拟结果见表 6.4.4. 表 6.4.4 还分别给出了变量选择过程平均计算时间. 从表 6.4.3

和表 6.4.4 可以得出以下结论:

表 6.4.3　基于不同的变量选择方法, 对参数分量的变量选择结果

	$n=100$			$n=150$			$n=200$		
方法	C	I	GMSE	C	I	GMSE	C	I	GMSE
sSCAD	6.667	0.003	0.015	6.773	0	0.012	6.879	0	0.007
pSCAD	5.780	0.001	0.039	6.357	0	0.018	6.805	0	0.011
Oracle	7	0	0.012	7	0	0.007	7	0	0.005

表 6.4.4　基于不同的变量选择方法, 对非参数分量的变量选择结果

	$n=100$				$n=200$			
方法	C	I	RASE	时间	C	I	RASE	时间
sSCAD	7.650	0	0.074	2.35	7.976	0	0.011	4.994
GLRT	7.646	0	0.076	4.51	7.974	0	0.017	15.728
Oracle	8	0	0.008	0.515	8	0	0.004	0.645

(i) 随着样本容量 n 的增大, 基于所有变量选择方法的结果均越来越接近于基于 oracle 方法所得的结果.

(ii) 对参数分量, 当样本容量 n 较大时, 基于 sSCAD 的变量选择结果类似于基于 pSCAD 的变量选择结果. 但是, 当 n 较小时, sSCAD 变量选择方法明显优于 pSCAD 变量选择方法.

(iii) 对非参数分量, 基于 sSCAD 方法的变量选择结果与基于 GLRT 方法的变量选择的结果类似. 但是, 利用 GLRT 方法的平均计算时间远远大于利用 sSCAD 方法所需的平均计算时间.

6.4.5　实例分析

下面应用 6.4.3 小节提出的变量选择方法对 5.4.4 小节所研究的 AIDS 病数据进行变量选择分析. 为了说明 6.4.3 小节所提出的方法, 本节忽略对每个个体重复测量所得数据的相关性.

令 Y 为在患者的 CD4 百分率, Z_1 为 HIV 感染前中心化的 CD4 百分率, Z_2 为患者的抽烟状态, 根据患者是否抽烟分别取值为 1 和 0, X 为 HIV 感染时患者的年龄. 考虑如下模型

$$Y = a_0(t) + a_1(t)X + \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon,$$

其中 $a_0(t)$ 为 CD4 百分率基准函数, 表示 HIV 感染 t 年后, 患者的平均 CD4 百分率.

利用 6.4.3 小节所提出的变量选择方法对上述模型进行分析, 得到两个非零的回归系数 $a_0(t)$ 和 β_1. 该结果表明患者的抽烟状态以及 HIV 感染时的年龄对 HIV

感染后患者的 CD4 百分率的影响不是太显著. 这一结论与 5.4.5 小节的结果基本上是一致的. 另外, 图 6.4.1 给出了基准函数 $a_0(t)$ 的估计曲线. 从图 6.4.1 可以看出, 关于基准函数 $a_0(t)$ 的估计曲线与 5.4.5 小节的结果是类似的.

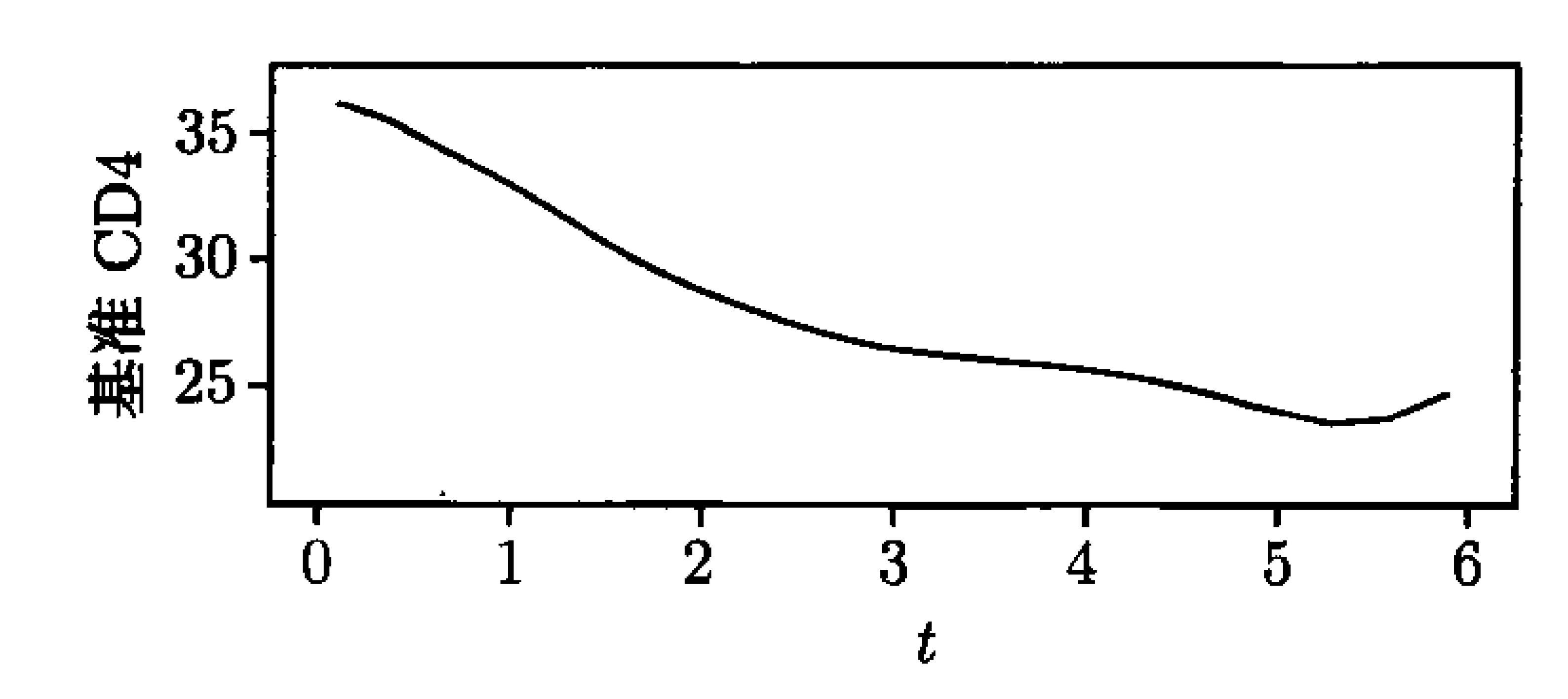


图 6.4.1 应用于 AIDS 病数据. 基于 sSCAD 方法所得基准函数 $a_0(t)$ 的正则估计

6.4.6 定理的证明

为书写方便, 下文用 C 表示某个正常数, 每次出现时可以代表不同的值.

定理 6.4.1 的证 令 $\delta = n^{-r/(2r+1)} + a_n$, $\beta = \beta_0 + \delta T_1$, $\gamma = \gamma_0 + \delta T_2$, $T = (T_1^{\mathrm{T}}, T_2^{\mathrm{T}})^{\mathrm{T}}$. 先考虑定理 6.4.1 的第一式. 首先证明, 对给定的 $\varepsilon > 0$, 则存在一个较大的 C 满足

$$P\left\{\inf_{||T||=C} Q(\gamma,\beta) > Q(\gamma_0,\beta_0)\right\} \geqslant 1-\varepsilon. \tag{6.4.10}$$

令 $\Delta(\gamma,\beta) = K^{-1}[Q(\gamma,\beta) - Q(\gamma_0,\beta_0)]$, 那么结合 $p_\lambda(0) = 0$, 并经简单计算可得

$$\begin{aligned}
\Delta(\gamma,\beta) \geqslant & -\frac{2\delta}{K}\sum_{i=1}^{n}[\varepsilon_i + X_i^{\mathrm{T}}R(U_i)][Z_i^{\mathrm{T}}T_1 + W_i^{\mathrm{T}}T_2] \\
& + \frac{\delta^2}{K}\sum_{i=1}^{n}[Z_i^{\mathrm{T}}T_1 + W_i^{\mathrm{T}}T_2]^2 + \frac{n}{K}\sum_{l=1}^{s}[p_{\lambda_{2l}}(|\beta_l|) - p_{\lambda_{2l}}(|\beta_{l_0}|)] \\
& + \frac{n}{K}\sum_{k=1}^{d}[p_{\lambda_{1k}}(||\gamma_k||_H) - p_{\lambda_{1k}}(||\gamma_{k_0}||_H)] \\
\equiv & \ S_1 + S_2 + S_3 + S_4,
\end{aligned}$$

其中 $R(u) = (R_1(u), \cdots, R_p(u))^{\mathrm{T}}$, $R_k(u) = \mathbf{a}_k(u) - \gamma_k^{\mathrm{T}}B(u)$, $k = 1, \cdots, p$. 由条件 (C6.4.2), (C6.4.3) 和 (Schumaker, 1981) 中的推论 6.21, 可得 $||R_k(u)|| = O(K^{-r})$. 于是, 再结合条件 (C6.4.2), 经简单计算可得

$$\sum_{i=1}^{n} X_i^{\mathrm{T}}R(U_i)[Z_i^{\mathrm{T}}T_1 + W_i^{\mathrm{T}}T_2] = O_p(nK^{-r}||T||). \tag{6.4.11}$$

注意到 $E[\varepsilon_i|Z_i, X_i, U_i] = 0$, 可以证明

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i[Z_i^{\mathrm{T}}T_1 + W_i^{\mathrm{T}}T_2] = O_p(||T||).$$

再由 (6.4.11) 式, 经简单计算可得

$$S_1 = O_p(\sqrt{n}K^{-1}\delta)||T|| + O_p(nK^{-1-r}\delta)||T|| = O_p(1 + n^{\frac{r}{2r+1}}a_n)||T||.$$

类似地, 可以证明

$$S_2 = O_p(nK^{-1}\delta^2)||T||^2 = O_p(1 + 2n^{\frac{r}{2r+1}}a_n)||T||^2.$$

因此, 对充分大的 C, 当 $||T|| = C$ 时, S_1 一致地被 S_2 控制. 利用 Taylor 展开可得

$$\begin{aligned} S_3 &= nK^{-1}\sum_{l=1}^{s}[p_{\lambda_{2l}}(|\beta_l|) - p_{\lambda_{2l}}(|\beta_{l_0}|)] \\ &\leqslant \sum_{l=1}^{s}\{nK^{-1}\delta p'_{\lambda_{2l}}(|\beta_{l0}|)\mathrm{sgn}(\beta_{l0})|T_{1l}| + nK^{-1}\delta^2 p''_{\lambda_{2l}}(|\beta_{l0}|)|T_{1l}|^2[1+o(1)]\} \\ &\leqslant \sqrt{s}K^{-1}n\delta a_n||T|| + nK^{-1}\delta^2 b_n||T||^2. \end{aligned}$$

因此, 当 $||T|| = C$ 时, S_3 一致地被 S_2 所控制. 利用同样的证明方法可以证明, 对充分大的 C, 当 $||T|| = C$ 时, I_4 一致地被 I_2 所控制. 因此, 通过选择充分大的 C, (6.4.10) 式成立. 这就说明, 在邻域 $\{\beta_0 + \delta T_1 : ||T_1|| \leqslant C\}$ 内, $Q(\gamma, \beta)$ 至少以概率 $1-\varepsilon$ 存在局部最小值. 因此, 存在满足 $||\hat{\gamma} - \gamma_0|| = O_p(\delta)$ 和 $||\hat{\beta} - \beta_0|| = O_p(\delta)$ 的 $\hat{\gamma}$ 和 $\hat{\beta}$, 使得 $Q(\gamma, \beta)$ 在点 $(\hat{\gamma}^{\mathrm{T}}, \hat{\beta}^{\mathrm{T}})^{\mathrm{T}}$ 处达到最小. 这就完成了定理 6.4.1 的第一式的证明.

下面证明定理 6.4.1 的第二式. 注意到

$$\begin{aligned} ||\hat{\boldsymbol{a}}_k(u) - \boldsymbol{a}_{k0}(u)||^2 &= \int_0^1 [\hat{\boldsymbol{a}}_k(u) - \boldsymbol{a}_{k0}(u)]^2 \mathrm{d}u \\ &= \int_0^1 [\hat{\gamma}_k^{\mathrm{T}} B(u) - \gamma_k^{\mathrm{T}} B(u) + R_k(u)]^2 \mathrm{d}u \\ &\leqslant 2\int_0^1 [\hat{\gamma}_k^{\mathrm{T}} B(u) - \gamma_k^{\mathrm{T}} B(u)]^2 \mathrm{d}u + 2\int_0^1 R_k^2(u)\mathrm{d}u \\ &= 2(\hat{\gamma}_k - \gamma_k)^{\mathrm{T}} H(\hat{\gamma}_k - \gamma_k) + 2\int_0^1 R_k^2(u)\mathrm{d}u. \end{aligned}$$

由第一式的证明过程可以得到 $||\hat{\gamma} - \gamma|| = O_p(n^{-r/(2r+1)} + a_n)$. 因此, 注意到 $||H|| = O(1)$ 并经简单计算可得

$$(\hat{\gamma}_k - \gamma_k)^{\mathrm{T}} H(\hat{\gamma}_k - \gamma_k) = O_p\left(n^{\frac{-2r}{2r+1}} + a_n^2\right). \tag{6.4.12}$$

另外, 经简单计算可知

$$\int_0^1 R_k^2(u)\mathrm{d}u = O_p\left(n^{\frac{-2r}{2r+1}}\right). \tag{6.4.13}$$

结合 (6.4.12) 和 (6.4.13) 式即可得到第二个结果的证明. 这就证完了定理 6.4.1 的证明.

定理 6.4.2 的证 首先证明第 (i) 部分. 由 $\lambda_{\max} \to 0$ 可知, 对充分大的 n, 有 $a_n = 0$. 因此, 根据定理 6.4.1, 只需证明, 对任意 γ 和 β_l, 且分别满足 $||\gamma - \gamma_0|| = O_p(n^{-r/(2r+1)})$ 和 $|\beta_l - \beta_{l0}| = O_p(n^{-r/(2r+1)})$, $l = 1, \cdots, s$, 以及对给定的 $\varepsilon = Cn^{-r/(2r+1)}$, 当 n 充分大时, 以概率 1 有

$$\frac{\partial Q(\gamma, \beta)}{\partial \beta_l} > 0, \quad 0 < \beta_l < \varepsilon, \ l = s+1, \cdots, q, \tag{6.4.14}$$

$$\frac{\partial Q(\gamma, \beta)}{\partial \beta_l} < 0, \quad -\varepsilon < \beta_l < 0, \ l = s+1, \cdots, q. \tag{6.4.15}$$

因此, (6.4.14) 和 (6.4.15) 式表明 $Q(\gamma, \beta)$ 的最小值在 $\beta_l = 0 (l = s+1, \cdots, q)$ 处达到.

利用与定理 6.4.1 类似的证明可得

$$\begin{aligned}
\frac{\partial Q(\gamma, \beta)}{\partial \beta_l} &= -2\sum_{i=1}^{n} Z_{il}(Y_i - Z_i^{\mathrm{T}}\beta - W_i^{\mathrm{T}}\gamma) + np'_{\lambda_{2l}}(|\beta_l|)\mathrm{sgn}(\beta_l) \\
&= -2\sum_{i=1}^{n} Z_{il}[\varepsilon_i + X_i^{\mathrm{T}}R(U_i)] - 2\sum_{i=1}^{n} Z_{il}Z_i^{\mathrm{T}}(\beta_0 - \beta) \\
&\quad -2\sum_{i=1}^{n} Z_{il}W_i^{\mathrm{T}}(\gamma_0 - \gamma) + np'_{\lambda_{2l}}(|\beta_l|)\mathrm{sgn}(\beta_l) \\
&= n\lambda_{2l}\left\{\lambda_{2l}^{-1}p'_{\lambda_{2l}}(|\beta_l|)\mathrm{sgn}(\beta_l) + O_p(\lambda_{2l}^{-1}n^{-\frac{r}{2r+1}})\right\},
\end{aligned}$$

其中 $\mathrm{sgn}(\cdot)$ 为符号函数. 注意到 $\lim\limits_{n\to\infty}\liminf\limits_{\beta_l\to 0}\lambda_{2l}^{-1}p'_{\lambda_{2l}}(|\beta_l|) > 0$ 和 $\lambda_{2l}n^{\frac{r}{2r+1}} > \lambda_{\min} n^{\frac{r}{2r+1}} \to \infty$, 因此, 上式的符号完全由 β_l 的符号决定, 即 (6.4.14) 和 (6.4.15) 式成立. 因此, 定理 6.4.2 的第 (i) 部分成立.

与证明第 (i) 部分完全类似的论证可知, 以概率 1 有 $\hat{\gamma}_k = 0$, $k = d+1, \cdots, p$. 结合 $\sup\limits_u ||B(u)|| = O(1)$ 和 $\hat{a}_k(u) = \hat{\gamma}_k^{\mathrm{T}}B(u)$ 即可证得定理 6.4.2 的第 (ii) 部分. 这就完成定理 6.4.2 的证明.

引理6.4.1 设条件 (C6.4.1)~(6.4.5) 成立, 且内部节点个数 $K = O(n^{1/(2r+1)})$, 则

$$\frac{1}{n}\sum_{i=1}^{n}\breve{Z}_i^*\breve{Z}_i^{*T} \xrightarrow{P} \Sigma^*,$$

其中 $\breve{Z}_i^* = Z_i^* - \Psi_n^{\mathrm{T}}\Phi_n^{-1}W_i^*$, $\Phi_n = n^{-1}\sum\limits_{i=1}^{n}W_i^*W_i^{*\mathrm{T}}$, $\Psi_n = n^{-1}\sum\limits_{i=1}^{n}W_i^*Z_i^{*\mathrm{T}}$.

证　记 $\boldsymbol{W}^* = (W_1^*, \cdots, W_n^*)^{\mathrm{T}}$, $\boldsymbol{Z}^* = (Z_1^*, \cdots, Z_n^*)^{\mathrm{T}}$ 以及 $\boldsymbol{Z}^* = (\boldsymbol{Z}^* - \Gamma_n) + \Gamma_n \equiv \Delta_n + \Gamma_n$, 其中 $\Gamma_n = (\Psi(U_1)\Phi(U_1)^{-1}X_1^*, \cdots, \Psi(U_n)\Phi(U_n)^{-1}X_n^*)^{\mathrm{T}}$. 经简单计算可得

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} \breve{Z}_i^* \breve{Z}_i^{*\mathrm{T}} &= n^{-1}\boldsymbol{Z}^{*\mathrm{T}}(I - P^{\mathrm{T}})(I - P)\boldsymbol{Z}^* \\ &= n^{-1}[\Delta_n^{\mathrm{T}}\Delta_n + \Gamma_n^{\mathrm{T}}(I - P^{\mathrm{T}})(I - P)\Gamma_n + \Delta_n^{\mathrm{T}}(I - P^{\mathrm{T}})(I - P)\Gamma_n \\ &\quad + \Gamma_n^{\mathrm{T}}(I - P^{\mathrm{T}})(I - P)\Delta_n + \Delta_n^{\mathrm{T}}P^{\mathrm{T}}P\Delta_n],\end{aligned} \tag{6.4.16}$$

其中 $P = \boldsymbol{W}^*(\boldsymbol{W}^{*\mathrm{T}}\boldsymbol{W}^*)^{-1}\boldsymbol{W}^{*\mathrm{T}}$. 由条件 (C6.4.2) 可知, 存在某一矩阵 M 使得 $||\Gamma_n - \boldsymbol{W}^*M|| = O_p(n^{1/2}K^{-r})$. 另外, 注意到 P 为投影矩阵, 有

$$\begin{aligned}||(I - P)\Gamma_n|| &= ||\Gamma_n - \boldsymbol{W}^*M|| + ||\boldsymbol{W}^*M - P\Gamma_n|| \\ &\leqslant 2||\Gamma_n - \boldsymbol{W}^*M|| = O_p(n^{1/2}K^{-r}).\end{aligned} \tag{6.4.17}$$

经计算可得 $E(\boldsymbol{W}^{*\mathrm{T}}\Delta_n|U_1, \cdots, U_n) = 0$. 因此, $E(\boldsymbol{W}^{*\mathrm{T}}\Delta_n) = 0$,

$$E(||\boldsymbol{W}^{*\mathrm{T}}\Delta_n||^2) = E\left(\sum_{i=1}^{n} ||W_i^*\Delta_{ni}||^2\right) = O_p(nK),$$

其中 Δ_{ni} 为 Δ_n 的第 i 行. 于是, $||\boldsymbol{W}^{*\mathrm{T}}\Delta_n|| = O_p(n^{1/2}K^{1/2})$. 从而可得

$$\begin{aligned}||P\Delta_n|| &\leqslant ||\boldsymbol{W}^*||||\boldsymbol{W}^{*\mathrm{T}}\boldsymbol{W}^*||||\boldsymbol{W}^{*\mathrm{T}}\Delta_n|| \\ &= O_p(n^{1/2}K^{-1/2})O_p(n^{-1}K)O_p(n^{1/2}K^{1/2}) = O_p(K),\end{aligned} \tag{6.4.18}$$

$$||(I - P)\Delta_n|| = O_p(n^{1/2}). \tag{6.4.19}$$

因此, 由 (6.4.17)~(6.4.19) 式可知, (6.4.16) 式右边除第一项外, 其他各项均为 $o_p(1)$. 进一步, 由大数定律可得, (6.4.16) 式右边第一项以概率收敛到 Σ^*, 这就完成了引理 6.4.1 的证明.

定理 6.4.3 的证　由定理 6.4.1 和定理 6.4.2 可知, 当 $n \to \infty$ 时, 以概率 1, $Q(\gamma, \beta)$ 在点 $(\hat{\beta}^{*\mathrm{T}}, 0)^{\mathrm{T}}$ 和 $(\hat{\gamma}^{*\mathrm{T}}, 0)^{\mathrm{T}}$ 处达到最小. 因此, 令 $Q_{1n}(\gamma, \beta) = \dfrac{\partial Q(\gamma, \beta)}{\partial \beta^*}$ 和 $Q_{2n}(\gamma, \beta) = \dfrac{\partial Q(\gamma, \beta)}{\partial \gamma^*}$, 那么, $(\hat{\beta}^{*\mathrm{T}}, 0)^{\mathrm{T}}$ 和 $(\hat{\gamma}^{*\mathrm{T}}, 0)^{\mathrm{T}}$ 一定满足

$$\begin{aligned}&n^{-1}Q_{1n}((\hat{\gamma}^{*\mathrm{T}}, 0)^{\mathrm{T}}, (\hat{\beta}^{*\mathrm{T}}, 0)^{\mathrm{T}}) \\ &= \frac{1}{n}\sum_{i=1}^{n} Z_i^*\{Y_i - Z_i^{*\mathrm{T}}\hat{\beta}^* - W_i^{*\mathrm{T}}\hat{\gamma}^*\} + \sum_{l=1}^{s} p'_{\lambda_{2l}}(|\hat{\beta}_l|)\mathrm{sgn}(\hat{\beta}_l) = 0.\end{aligned} \tag{6.4.20}$$

$$\begin{aligned}&n^{-1}Q_{2n}((\hat{\gamma}^{*\mathrm{T}},0)^{\mathrm{T}},(\hat{\beta}^{*\mathrm{T}},0)^{\mathrm{T}})\\&=\frac{1}{n}\sum_{i=1}^{n}W_i^*\{Y_i-Z_i^{*\mathrm{T}}\hat{\beta}^*-W_i^{*\mathrm{T}}\hat{\gamma}^*\}+\sum_{k=1}^{d}p'_{\lambda_{1k}}(\|\hat{\gamma}_k\|_H)\frac{H\hat{\gamma}_k}{\|\hat{\gamma}_k\|_H}=0.\end{aligned}\tag{6.4.21}$$

对 $p'_{\lambda_{2l}}(|\hat{\beta}_l|)$ 应用 Taylor 展开可得

$$p'_{\lambda_{2l}}(|\hat{\beta}_l|)=p'_{\lambda_{2l}}(|\beta_{l_0}|)+[p''_{\lambda_{2l}}(|\beta_{l_0}|)+o_p(1)](\hat{\beta}_l-\beta_{l_0}).$$

另外, 条件 (C6.4.4) 表明 $p''_{\lambda_{2l}}(|\beta_{l0}|)=o_p(1)$. 并且注意到当 $\lambda_{\max}\to 0$ 时, $p'_{\lambda_{2l}}(|\beta_{l0}|)=0$, 从而有 $\sum_{l=1}^{s}p'_{\lambda_{2l}}(|\hat{\beta}_l|)\mathrm{sgn}(\hat{\beta}_l)=o_p(\hat{\beta}^*-\beta_0^*)$. 因此, 结合 (6.4.20) 式, 经简单计算可得

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}Z_i^*[Z_i^{*\mathrm{T}}(\beta_0^*-\hat{\beta}^*)+W_i^{*\mathrm{T}}(\gamma_0^*-\hat{\gamma}^*)+X_i^{*\mathrm{T}}R^*(U_i)+\varepsilon_i]\\&\quad+o_p(\hat{\beta}^*-\beta_0^*)=0,\end{aligned}\tag{6.4.22}$$

其中 $X_i^*=(X_{i1},\cdots,X_{id})^{\mathrm{T}}$ 以及 $R^*(u)=(R_1(u),\cdots,R_d(u))^{\mathrm{T}}$. 结合 (6.4.21) 式, 并利用与证明 (6.4.22) 式类似的方法, 可以证明

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}W_i^*[Z_i^{*\mathrm{T}}(\beta_0^*-\hat{\beta}^*)+W_i^{*\mathrm{T}}(\gamma_0^*-\hat{\gamma}^*)+X_i^{*\mathrm{T}}R^*(U_i)+\varepsilon_i]\\&\quad+o_p(\hat{\gamma}^*-\gamma_0^*)=0.\end{aligned}\tag{6.4.23}$$

从而, 由 (6.4.23) 式可得

$$\hat{\gamma}^*-\gamma_0^*=[\Phi_n+o_p(1)]^{-1}\left[\Psi_n(\beta_0^*-\hat{\beta}^*)+\frac{1}{n}\sum_{i=1}^{n}W_i^*[X_i^{*\mathrm{T}}R^*(U_i)+\varepsilon_i]\right].$$

把上式代入 (6.4.22) 式, 并经简单计算可得

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}Z_i^*[Z_i^*-\Psi_n^{\mathrm{T}}\Phi_n^{-1}W_i^*]^{\mathrm{T}}(\hat{\beta}^*-\beta_0^*)+o_p(\hat{\beta}^*-\beta_0^*)\\&=\frac{1}{n}\sum_{i=1}^{n}Z_i^*\left\{\varepsilon_i+X_i^{*\mathrm{T}}R^*(U_i)-W_i^{*\mathrm{T}}[\Phi_n^{-1}+o_p(1)]\Lambda_n\right\},\end{aligned}\tag{6.4.24}$$

其中 $\Lambda_n=n^{-1}\sum_{j=1}^{n}W_j^*[X_j^{*\mathrm{T}}R^*(U_j)+\varepsilon_j]$. 注意到

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_n^{\mathrm{T}}\Phi_n^{-1}W_i^*[Z_i^{*\mathrm{T}}-W_i^{*\mathrm{T}}\Phi_n^{-1}\Psi_n]=0,$$

以及

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_n^{\mathrm{T}}\Phi_n^{-1}W_i^*\left[\varepsilon_i+X_i^{*\mathrm{T}}R^*(U_i)-W_i^{*\mathrm{T}}\Phi_n^{-1}\Lambda_n\right]=0.$$

那么, 由 (6.4.24) 式, 容易证明

$$\begin{aligned}&\left[\frac{1}{n}\sum_{i=1}^{n}\breve{Z}_i^*\breve{Z}_i^{*\mathrm{T}}+o_p(1)\right]\sqrt{n}(\hat{\beta}^*-\beta_0^*)\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\breve{Z}_i^*\varepsilon_i+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\breve{Z}_i^*W_i^{*\mathrm{T}}[\Phi_n^{-1}+o_p(1)]\Lambda_n+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\breve{Z}_i^*X_i^{*\mathrm{T}}R^*(U_i)\\&\equiv B_1+B_2+B_3,\end{aligned}\tag{6.4.25}$$

其中 $\breve{Z}_i^*=Z_i^*-\Psi_n^{\mathrm{T}}\Phi_n^{-1}W_i^*$. 类似引理 6.4.1 的证明可以证得

$$\frac{1}{n}\sum_{i=1}^{n}\breve{Z}_i^*\breve{Z}_i^{*\mathrm{T}}\varepsilon_i^2\xrightarrow{P}\sigma^2\Sigma^*.$$

从而, 利用中心极定理可得

$$B_1\xrightarrow{D}N(0,\sigma^2\Sigma^*).\tag{6.4.26}$$

另外, 注意到 $\sum_{i=1}^{n}\breve{Z}_i^*W_i^{*\mathrm{T}}=0$, 我们有 $B_2=0$. 经简单计算得

$$\begin{aligned}B_3&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Z_i^*-E(\Psi_n)^{\mathrm{T}}E(\Phi_n)^{-1}W_i^*]X_i^{*\mathrm{T}}R^*(U_i)\\&=+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[E(\Psi_n)^{\mathrm{T}}E(\Phi_n)^{-1}-\Psi_n^{\mathrm{T}}\Phi_n^{-1}]W_i^*X_i^{*\mathrm{T}}R^*(U_i)\\&\equiv B_{31}+B_{32}.\end{aligned}$$

注意到 $E\{[Z_i^*-E(\Psi_n)^{\mathrm{T}}E(\Phi_n)^{-1}W_i^*]W_i^{*\mathrm{T}}\}=0$, 可以证明

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Z_i^*-E(\Psi_n)^{\mathrm{T}}E(\Phi_n)^{-1}W_i^*]W_i^{*\mathrm{T}}=O_p(1).$$

再结合 $W_i=I_p\otimes B(U_i)X_i$, $||B(u)||=O(1)$ 以及 $||R(u)||=o(1)$, 显然有 $B_{31}=o_p(1)$. 类似地, 可以证明 $B_{32}=o_p(1)$. 因此, $B_3=o_p(1)$. 最后, 结合 (6.4.25) 和 (6.4.26) 式以及引理 6.4.1, 并利用 Slutsky 定理即可完成定理 6.4.3 的证明.

参 考 文 献

Ahmad I, Leelahanon S, Li Q. 2005. Efficient estimation of a semiparametric partially linear varying coefficient model. The Annals of Statistics, 33(1): 258~283

Bickel P J, Kwon J. 2001. Inference for semiparametric models: Some current frontiers (with discussion). Statistica Sinica, 11: 863~960

Chamberlain G. 1992. Efficiency bounds for semiparametric regression. Econometrica, 60: 567～596

Fan J, Huang T. 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli, 11(6): 1031～1057

Fan J, Li R Z. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96: 1348～1360

Hu T, Cui H J. 2010. Robust estimates in generalised varying-coefficient partially linear models. Journal of Nonparametric Statistics, 22(6): 737～754

Hu X M, Wang Z Z, Liu F. 2009. Testing serial correlation in semiparametric varying-coefficient partially linear models. Communications in Statistics-Theory and Methods, 38(13): 2145～2163

Huang Z S. 2011. Empirical likelihood for generalized partially linear varying-coefficient models. Journal of Applied Statistics, 38(6): 1265～1275

Huang Z S, Zhang R Q. 2009. Empirical likelihood for nonparametric parts in semiparametric varying-coefficient partially linear models. Statistics & Probability Letters, 79(16): 1798～1808

Huang Z S, Zhou Z G, Jiang R, 2010. Empirical likelihood based inference for semiparametric varying coefficient partially linear models with error-prone linear covariates. Statistics & Probability Letters, 80(5-6): 497～504

Kai B, Li R, Zou H. 2011. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. The Annals of Statistics, 39(1): 305～332

Lam C, Fan J. 2008. Profile-kernel likelihood inference with diverging number of parameters. The Annals of Statistics, 36(5): 2232～2260

Li D G, Chen J, Lin Z Y. 2009. Variable selection in partially time-varying coefficient models. Journal of Nonparametric Statistics, 21(5): 553～566

Li D G, Chen J, Lin Z Y. 2011. Statistical inference in partially time-varying coefficient models. Journal of Statistical Planning and Inference, 141(2): 995～1013

Li G R, Feng S Y, Peng H. 2011. A profile-type smoothed score function for a varying coefficient partially linear model. Journal of Multivariate Analysis, 102(2): 372～385

Li G R, Xue L G, Lian H. 2011. Semi-varying coefficient models with a diverging number of components. Journal of Multivariate Analysis, 102(7): 1166～1174

Li R Z, Liang H. 2008. Variable selection in semiparametric regression modeling. The Annals of Statistics, 36: 261～286

Lia J L, Palta M. 2009. Bandwidth selection through cross-validation for semi-parametric varying-coefficient partially linear models. Journal of Statistical Computation and Simulation, 79(11): 1277～1286

Schumaker L L. 1981. Spline Functions, New York: Wiley

Stone C J. 1982. Optimal global rates of convergence for nonparametric regression. The Annals of Statistics, 10: 1348～1360

Wang J H, Zhu Z, Zhou J. 2009. Quantile regression in partially linear varying coefficient models. The Annals of Statistics. 37: 3841～3866

Wang X L, Li G R, Lin L. 2011. Empirical likelihood inference for semi-parametric varying-coefficient partially linear EV models. Metrika, 73(2): 171～185

Xia Y, Zhang W, Tong H. 2004. Efficient estimation for semivarying-coefficient models. Biometrika, 91(3): 661～681

Yang H, Li T T. 2010. Empirical likelihood for semiparametric varying coefficient partially linear models with longitudinal data. Statistics & Probability Letters, 80(2): 111～121

You J H, Chen G M. 2006. Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model. Journal of Multivariate Analysis, 97(2): 324～341

You J H, Zhou Y. 2006. Empirical likelihood for semiparametric varying-coefficient partially linear regression models. Statistics & Probability Letters, 76(4): 412～422

Zhang W. 2002. Local polynomial fitting in semivarying coefficient model. Journal of Multivariate Analysis, 82: 166～188

Zhang W W, Li G R, Xue L G. 2011. Profile inference on partially linear varying-coefficient errors-in-variables models under restricted condition. Computational Statistics & Data Analysis, 55(11): 3027～3040

Zhao P X, Xue L G. 2009a. Variable selection for semi-parametric varying coefficient partially linear models. Statistics & Probability Letters, 79: 2148～2157

Zhao P X, Xue L G. 2009b. Empirical likelihood inferences for semiparametric varying-coefficient partially linear errors-in-variables models with longitudinal data. Journal of Nonparametric Statistics, 21: 907～923

Zhao P X, Xue L G. 2010a. Variable selection for semiparametric varying coefficient partially linear errors-in-variables models. Journal of Multivariate Analysis, 101: 1872～1883

Zhao P X, Xue L G. 2010b. Empirical likelihood inferences for semiparametric varying-coefficient partially linear models with longitudinal data. Communications in Statistics-Theory and Methods, 39: 1898～1914

Zhou X, You J H. 2004. Wavelet estimation in varying-coefficient partially linear regression models. Statistics & Probability Letters, 68(1): 91～104

Zhou Y, Liang H. 2009. Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. The Annals of Statistics, 37(1): 427～458

Zhou Y, You J H, Wang X J. 2009. Strong convergence rates of several estimators in semiparametric varying-coefficient partially linear models. Acta Mathematica Scientia, 29(5): 1113～1127

第 7 章　单指标变系数模型

7.1　引　　言

考虑单指标变系数模型

$$Y = \boldsymbol{g}_0^{\mathrm{T}}(\beta_0^{\mathrm{T}} X) Z + \varepsilon, \tag{7.1.1}$$

其中 $(X, Z) \in \mathbf{R}^p \times \mathbf{R}^q$ 是响应变量 Y 的协变量, β_0 是 $p \times 1$ 未知参数向量, $g_0(\cdot)$ 是 $q \times 1$ 未知函数向量, 模型误差 ε 具有均值 0 和方差 σ^2, 且与 (X, Z) 独立. 为了可识别性, 假定 $||\beta_0|| = 1$ 且 β_0 的第一个非零分量是正的, 其中 $||\cdot||$ 表示 Euclid 模. 通常, Z 的第一个分量可以是 1.

模型 (7.1.1) 的主要优点是它避免了多维参数中经常遇到的维数灾祸, 其原因是 $g_0(\cdot)$ 是单变量函数向量. 模型 (7.1.1) 包括一类重要的统计模型. 若 $q = 1$ 和 $Z = 1$, 则 (7.1.1) 简化成单指标模型, 此模型由许多学者所研究, 参看第 3 章. 若 $p = 1$ 和 $\beta_0 = 1$, 则 (7.1.1) 即为变系数模型, 参看第 5 章. 若 β_0 的第 p 个分量非零, 令 $Z = (1, \breve{X}^{\mathrm{T}})^{\mathrm{T}}$, 其中 $\breve{X}$ 为去掉第 p 个分量后剩余的向量, 则 (7.1.1) 成为自适应变系数线性模型, Fan 等 (2003) 和 Lu 等 (2007) 研究了此模型的估计问题. Xia 和 Li(1999) 考虑了模型 (7.1.1) 的一般形式. Xue 和 Wang(2012) 使用经验似然方法研究了模型 (7.1.1) 中 β_0 的置信域的构造问题. 其他研究工作见 (Wu et al., 2011) 等.

通常使用单指标模型是因为它作为参数模型和非参数模型的合理折中而出现, 但依赖于参数模型的假定太强. 显然, 引入变系数模型部分推广了这个模型. 因此, 模型 (7.1.1) 在实际应用中容易解释, 因为它兼有单指标模型和变系数模型的特征.

本章的目的是阐述参数 β_0 的估计理论和方法, 其中包括点估计和区间估计. 然后给出估计量的渐近性质和置信域的构造. 需要指出的是, 本章的主要内容取材于 (Xue and Wang, 2012; Xue and Pang, 2012).

7.2　估计方法和主要结果

7.2.1　估计方法

假设 $\{(Y_i, X_i, Z_i), 1 \leqslant i \leqslant n\}$ 是来自模型 (7.1.1) 的样本, 那么

$$Y_i = \boldsymbol{g}_0^{\mathrm{T}}(\beta_0^{\mathrm{T}} X_i) Z_i + \varepsilon_i, \quad i = 1, \cdots, n, \tag{7.2.1}$$

其中 $X_i=(X_{i1},\cdots,X_{ip})^{\mathrm{T}}$, $Z_i=(Z_{i1},\cdots,Z_{iq})^{\mathrm{T}}$, ε_i 是均值 0 且方差为 σ^2 的随机误差. 假设 $\{\varepsilon_i,1\leqslant i\leqslant n\}$ 与 $\{(X_i,Z_i),1\leqslant i\leqslant n\}$ 独立.

记 $\mathcal{B}=\{\beta\in\mathbf{R}^p:||\beta||=1$, 且第一个非零元素是正的$\}$. 那么 β_0 是集合 $\mathcal{B}$ 的内点. 因此, 我们仅需在 $\mathcal{B}$ 上搜寻 β_0. 如果 $g_0(\cdot)$ 已知, 那么极小化

$$Q(\beta)\equiv E\left[Y-\boldsymbol{g}_0^{\mathrm{T}}(\beta^{\mathrm{T}}X)Z\right]^2,\quad ||\beta||=1 \tag{7.2.2}$$

可得到单指标方向 β_0. 如果使用 Newton 算法找 $Q(\beta)$ 的极小值点, 我们不需在点 β 处计算 $Q(\beta)$ 的导数. 然而, $\boldsymbol{g}_0(\beta^{\mathrm{T}}X)$ 的每个分量在点 β 处没有导数, 因为 $||\beta||=1$ 意味着 β 的真值是单位球上的边界点. 为此, 我们建议再参数化 β, 以便我们可以在 Euclid 空间 $\mathbf{R}^{p-1}$ 的一个区域上搜寻方向 β_0. 可以对 β_0 应用通常的 "去一分量" 的方法; 例如, Yu 和 Ruppert(2002) 与 Wang 等 (2010). 不失一般性, 可以假定真参数 β_0 的第 r 个分量是正的. 对 $\beta=(\beta_1,\cdots,\beta_p)^{\mathrm{T}}$, 设 $\beta^{(r)}=(\beta_1,\cdots,\beta_{r-1},\beta_{r+1},\cdots,\beta_p)^{\mathrm{T}}$ 是去掉 β 的第 r 个分量 β_r 后的 $p-1$ 维参数向量. 则新参数 $\beta_0^{(r)}$ 一定满足约束 $||\beta_0^{(r)}||<1$, 因此 β 在 $\beta_0^{(r)}$ 的邻域内是无限可微的. 注意到 $\beta_r=(1-||\beta^{(r)}||^2)^{1/2}$, 可得到 β 关于 $\beta^{(r)}$ 的 Jacobian 矩阵

$$J_{\beta^{(r)}}=\frac{\partial\beta}{\partial\beta^{(r)}}=(\gamma_1,\cdots,\gamma_p)^{\mathrm{T}}, \tag{7.2.3}$$

其中 $\gamma_s(1\leqslant s\leqslant p,s\neq r)$ 是 $p-1$ 维单位向量其第 $s-I(s>\gamma)$ 个分量是 1, 且 $\gamma_r=-(1-||\beta^{(r)}||^2)^{-1/2}\beta^{(r)}$.

现在考虑最小化 (7.2.2) 式中的 $Q(\beta)$ 的问题. 显然最小化 $Q(\beta)$ 等价于解估计方程

$$\left\{\begin{array}{l}E\{[Y-\boldsymbol{g}_0^{\mathrm{T}}(\beta^{\mathrm{T}}X)Z]\dot{\boldsymbol{g}}_0^{\mathrm{T}}(\beta^{\mathrm{T}}X)ZJ_{\beta^{(r)}}^{\mathrm{T}}X\}=0,\\ ||\beta||-1=0,\end{array}\right. \tag{7.2.4}$$

其中 $\dot{\boldsymbol{g}}_0(\cdot)$ 表示函数向量 $\boldsymbol{g}_0(\cdot)$ 的导数. 因为 $\boldsymbol{g}_0(\cdot)$ 和 $\dot{\boldsymbol{g}}_0(\cdot)$ 未知的, 我们需要分别用它们的估计量代替它们, 然后得到 (7.2.4) 式的样本形式. 用 $g_j(\cdot)$ 表示 $\boldsymbol{g}_0(\cdot)$ 的第 j 个分量. 现在使用局部线性拟合估计函数 $\{g_j(\cdot);j=1,\cdots,q\}$(Fan and Gijbels, 1996). 对 u 的一个小邻域 $\mathcal{U}$, 可以用下列线性函数局部地逼近 $g_j(U)$:

$$g_j(U)\approx g_j(u)+\dot{g}_j(u)(U-u)\equiv a_j+b_j(U-u),\quad j=1,\cdots,q.$$

$g_j(u)$ 和 $\dot{g}_j(u)$ 在点 β_0 处的局部线性估计定义为 $\hat{g}_j(u;\beta_0)=\hat{a}_j$ 和 $\hat{\dot{g}}_j(u;\beta_0)=\hat{b}_j$, 其中 $\{(\hat{a}_j,\hat{b}_j);j=1,\cdots,q\}$ 是下列加权和的极小值:

$$\sum_{i=1}^{n}\left\{Y_i-\sum_{j=1}^{q}\left[a_j+b_j(\beta_0^{\mathrm{T}}X_i-u)\right]Z_{ij}\right\}^2K_h(\beta_0^{\mathrm{T}}X_i-u),$$

其中 $K_h(\cdot)=h^{-1}K(\cdot/h)$, $K(\cdot)$ 是核函数, $h=h_n$ 是一列递减到 0 的带宽. 由最小二乘理论可以推出

$$\left(\hat{\boldsymbol{g}}^{\mathrm{T}}(u;\beta_0), h\hat{\dot{\boldsymbol{g}}}^{\mathrm{T}}(u;\beta_0)\right)^{\mathrm{T}} = S_n^{-1}(u;\beta_0)\xi_n(u;\beta_0), \tag{7.2.5}$$

其中 $\hat{\boldsymbol{g}}(u;\beta_0)=(\hat{g}_1(u;\beta_0),\cdots,\hat{g}_q(u;\beta_0))^{\mathrm{T}}$, $\hat{\dot{\boldsymbol{g}}}(u;\beta_0)=(\hat{\dot{g}}_1(u;\beta_0),\cdots,\hat{\dot{g}}_q(u;\beta_0))^{\mathrm{T}}$,

$$S_n(u;\beta_0)=\begin{pmatrix} S_{n,0}(u;\beta_0) & S_{n,1}(u;\beta_0) \\ S_{n,1}(u;\beta_0) & S_{n,2}(u;\beta_0)\end{pmatrix},$$

$$\xi_n(u;\beta_0)=\begin{pmatrix}\xi_{n,0}(u;\beta_0)\\ \xi_{n,1}(u;\beta_0)\end{pmatrix},$$

$$S_{n,j}(u;\beta_0)=\frac{1}{n}\sum_{i=1}^{n}Z_iZ_i^{\mathrm{T}}\left(\frac{\beta_0^{\mathrm{T}}X_i-u}{h}\right)^j K_h(\beta_0^{\mathrm{T}}X_i-u), \tag{7.2.6}$$

$$\xi_{n,j}(u;\beta_0)=\frac{1}{n}\sum_{i=1}^{n}Z_iY_i\left(\frac{\beta_0^{\mathrm{T}}X_i-u}{h}\right)^j K_h(\beta_0^{\mathrm{T}}X_i-u). \tag{7.2.7}$$

由于 $\boldsymbol{g}'(u)$ 的估计量的收敛速度比 $\boldsymbol{g}(u)$ 的估计量低, 因此如果使用同一个带宽，就会导致估计量 $\hat{\beta}$ 比 $\sqrt{n}$ 低的收敛速度. 为提高 $\boldsymbol{g}'(u)$ 的估计量的收敛速度, 我们在 $\hat{\dot{\boldsymbol{g}}}(u;\beta)$ 中引入另一个带宽 h_1 代替 h, 并定义为 $\hat{\dot{\boldsymbol{g}}}_{h_1}(u;\beta)$.

β_0 的一个估计量 $\hat{\beta}$ 定义为下列估计方程的解:

$$\begin{cases}\displaystyle\frac{1}{n}\sum_{i=1}^{n}[Y_i-\hat{\boldsymbol{g}}^{\mathrm{T}}(\beta^{\mathrm{T}}X_i;\beta)Z_i]\hat{\dot{\boldsymbol{g}}}_{h_1}^{\mathrm{T}}(\beta^{\mathrm{T}}X_i;\beta)Z_iJ_{\beta^{(r)}}^{\mathrm{T}}X_iw(\beta^{\mathrm{T}}X_i)=0,\\ ||\beta||-1=0,\end{cases} \tag{7.2.8}$$

其中 $w(\cdot)$ 是具有有界支撑 $\mathcal{U}_w$ 的有界权函数, 它是用来控制 $g_0(\cdot)$ 和 $\dot{g}_0(\cdot)$ 的估计量的边界效应. 这里取示性函数 $w(\cdot)=I_{\mathcal{U}_w}(\cdot)$.

利用 $\hat{\beta}$, 我们定义 $\boldsymbol{g}(u)$ 的估计量为 $\hat{\boldsymbol{g}}^*(u)=\hat{\boldsymbol{g}}(u,\hat{\beta})$, σ^2 的估计量为

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}[Y_i-\hat{\boldsymbol{g}}^{\mathrm{T}}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})Z_i]^2. \tag{7.2.9}$$

7.2.2 带宽选择

诸如交叉验证和广义交叉验证的各种带宽选择方法都适合于估计量 $\hat{g}$. 在下面的模拟中, 我们使用 Cai 等 (2000) 提出的多折叠交叉验证 (MCV) 准则来选择最优带宽, 因为这个算法既简单又快速. 设 m 和 Q 是两个给定的正整数且 $n>mQ$. 基本思想是首先使用长度为 $n-km(k=1,\cdots,Q)$ 的 Q 个子序列估计未知系数函数,

然后基于估计的模型计算长度为 m 的一步预测误差. 更确切地, 我们选择 h 使得下式极小化:

$$\mathrm{AMS}(h) = \sum_{k=1}^{Q} \mathrm{AMS}_k(h), \tag{7.2.10}$$

其中

$$\mathrm{AMS}_k(h) = \frac{1}{m} \sum_{i=n-km+1}^{n-km+m} \left[Y_i - \sum_{j=1}^{q} \hat{g}^*_{j,k}(U_i) Z_{ij} \right]^2, \quad k = 1, \cdots, Q,$$

$\hat{g}^*_{j,k}(\cdot)$ 是 $\boldsymbol{g}_0(\cdot)$ 的第 j 个分量的估计量, 它是利用样本 $\{(Y_i, U_i, Z_i), 1 \leqslant i \leqslant n - km\}$ 来计算, 其带宽为 $h\left(\dfrac{n}{n-km}\right)^{1/5}$. 这个带宽满足 $h\left(\dfrac{n}{n-km}\right)^{1/5} \propto (n-km)^{-1/5}$.

注意到对不同的样本量, 按照它的最优速度重置带宽, 即 $h \propto n^{-1/5}$. 因为选择的带宽不直接依赖于 m 和 Q, 为计算方便, 在下面的模拟中取 $m = [0.1n]$ 和 $Q = 4$.

设 h_{opt} 是极小化 (7.2.10) 得到的带宽, 即 $h_{\mathrm{opt}} = \inf\limits_{h>0} \mathrm{AMS}(h)$. 那么 h_{opt} 是估计 $\hat{g}(\cdot)$ 的最优带宽. 当计算估计量 $\hat{\beta}$, 使用近似带宽

$$\hat{h} = h_{\mathrm{opt}} n^{-1/20} (\log n)^{-1/2}, \quad \hat{h}_1 = h_{\mathrm{opt}}, \tag{7.2.11}$$

因为这就确保了对最优的渐近性质所需要的带宽有正确的量级 (Carroll et al., 1997), 并且带宽 $\hat{h}$ 和 $\hat{h}_1$ 满足条件 (C7.2.4).

7.2.3 算法

现在概述对 β_0 和 $\boldsymbol{g}_0(\cdot)$ 估计的算法.

步骤 0 (初始步): 指定 β 的模等于 1 的一个初始值.

步骤 1: 对给定的 β, 用公式 (7.2.5) 估计 $\boldsymbol{g}_0(\cdot)$ 和 $\dot{\boldsymbol{g}}_0(\cdot)$.

步骤 2: 对给定的 $\boldsymbol{g}_0(\cdot)$ 和 $\dot{\boldsymbol{g}}_0(\cdot)$, 使用估计方程 (7.2.8) 搜寻 β.

步骤 3: 重复步骤 1 和 2 直到收敛.

步骤 4: 利用由步骤 3 得到的估计值 $\hat{\beta}$, 得到 $\boldsymbol{g}_0(\cdot)$ 的最终估计为 $\hat{\boldsymbol{g}}^*(u) = \hat{\boldsymbol{g}}(u, \hat{\beta})$.

前述算法的基本思想很简单: 首先利用 (7.2.5) 式局部地估计 $\boldsymbol{g}_0(\cdot)$ 和 $\dot{\boldsymbol{g}}_0(\cdot)$, 然后使用所有数据和 (7.2.8) 式估计 β_0.

上述估计步骤涉及选择两个完全不同标准的光滑参数. 在算法的步骤 1 和 2 中, 目的是估计参数分量 β_0, 因此带宽 $\hat{h}$ 和 $\hat{h}_1$ 对这个执行是最优的. 然而, 在步骤 4 中, 目的是估计非参数分量 $\boldsymbol{g}_0(\cdot)$, 因此带宽 h_{opt} 对该步是最优的.

常常注意到非线性最优的收敛速度对初始值是敏感的. 对某些联系函数, 我们可以使用广义线性模型得到 β_0 的初始值. 但有时候就不能使用广义线性模型, 如

联系函数为指数函数和三角函数. 也注意到当指标系数的维数比较大时, 收敛速度可能很慢. 初始值的选择是一个公开的问题.

7.2.4 主要结果

为得到估计量的渐近性质, 我们首先给出下列条件

(C7.2.1) $\beta^{\mathrm{T}}X$ 的密度函数 $f(u)$ 对 $u \in \mathcal{U}_w$ 和接近于 β_0 的点 β 有界远离 0, 且在 $\mathcal{U}_w$ 上满足 1 阶 Lipschitz 条件, 其中 $\mathcal{U}_w$ 是 $w(u)$ 的支撑.

(C7.2.2) 函数 $g_j(u)$ $(1 \leqslant j \leqslant q)$ 在 $\mathcal{U}_w$ 上具有有界连续的二阶导数, 其中 $g_j(u)$ 是 $g_0(u)$ 的第 j 个分量.

(C7.2.3) $E(\|X\|^6) < \infty$, $E(\|Z\|^6) < \infty$, $E(\varepsilon^6) < \infty$.

(C7.2.4) 核 $K(u)$ 在其支撑 $(-1,1)$ 上是具有有界导数的对称概率密度函数.

(C7.2.5) $nh^2/\log^2 n \to \infty$, $nh^4 \log n \to 0$; $nhh_1^3/\log^2 n \to \infty$, $nh_1^5 = O(1)$.

(C7.2.6) 矩阵 $D(u)$ 是正定的, 且 $D(u)$ 和 $C(u)$ 的每一个元素在 $\mathcal{U}_w$ 上满足 1 阶 Lipschitz 条件, 其中 $D(u) = E(ZZ^{\mathrm{T}}|\beta_0^{\mathrm{T}}X = u)$, $C(u) = E(VZ^{\mathrm{T}}|\beta_0^{\mathrm{T}}X = u)$, $V = J_{\beta_0^{(r)}}^{\mathrm{T}} X \dot{\boldsymbol{g}}_0^{\mathrm{T}}(\beta_0^{\mathrm{T}}X) Z w(\beta_0^{\mathrm{T}}X)$, $\mathcal{U}_w$ 在 (C7.2.1) 中定义.

(C7.2.7) $\{(Y_i, X_i, Z_i), 1 \leqslant i \leqslant n\}$ 是严平稳强混合序列, 具有混合系数 $\alpha(n) = O(\rho^n)$, 其中 $0 < \rho < 1$ 为某个常数.

(C7.2.8) 条件密度 $f_{U_1,U_l|Y_1,Y_l,Z_1,Z_l}(u_1, u_l|y_1, y_l, z_1, z_l)$ 对任何 $l > 1$ 是有界的, 其中 $U_i = \beta_0^{\mathrm{T}}X_i$, $i \geqslant 1$.

对某个正的常数 c_1, 记 $\mathcal{B}_n = \{\beta \in \mathcal{B} : \|\beta - \beta_0\| \leqslant c_1 n^{-1/2}\}$. 下列引理需要约束 $\beta \in \mathcal{B}_n$. 该约束受下面事实启发: 因为我们期望 $\hat{\beta}$ 是 $\sqrt{n}$ 相合的, 我们将从距 β_0 的阶为 $n^{-1/2}$ 的 β 中寻找方程 (7.2.8) 的解. 从定理 7.2.1 可以看到 $\|\hat{\beta} - \beta_0\| = O_P(n^{-1/2})$, 于是, $\hat{\beta} \in \mathcal{B}_n$. 因此, 将 β 约束在圆锥体 $\mathcal{B}_n$ 内并没有排除任何感兴趣的最小值, 也没有失去一般性. Härdle 等 (1993) 与 Xia 和 Li(1999) 也作了类似的约束. 提出这个范围仅仅是为了理论证明. 实际上, 我们仅需假定 $Q(\beta)$ 在 $\mathcal{B}$ 内有唯一的最小值点. 记 $\mu_j = \int u^j K(u)\mathrm{d}u$, $\nu_j = \int u^j K^2(u)\mathrm{d}u$, $j = 1, 2, 3$. 我们首先陈述估计量 $\hat{\boldsymbol{g}}(\cdot)$ 和 $\hat{\dot{\boldsymbol{g}}}(\cdot)$ 的一致收敛速度.

定理 7.2.1 设条件 (C7.2.1)~(C7.2.4) 和 (C7.2.6)~(C7.2.8) 成立, 则

$$\sup_{u \in \mathcal{U}_w, \beta \in \mathcal{B}_n} \|\hat{\boldsymbol{g}}(u;\beta) - \boldsymbol{g}_0(u)\| = O\left(\left\{\frac{\log n}{nh}\right\}^{1/2} + h^2\right), \quad \text{a.s.},$$

$$\sup_{u \in \mathcal{U}_w, \beta \in \mathcal{B}_n} \|\hat{\dot{\boldsymbol{g}}}(u;\beta) - \dot{\boldsymbol{g}}_0(u)\| = O\left(\left\{\frac{\log n}{nh^3}\right\}^{1/2} + h\right), \quad \text{a.s.}$$

由 Stute(1982) 的结果和重对数律, 我们可以看到前述 $\hat{\boldsymbol{g}}(u;\beta)$ 的收敛速度是最优的. 下列定理陈述了 $\hat{\beta}$ 的渐近正态性.

定理 7.2.2 设条件 (C7.2.1)~(C7.2.8) 成立, 则

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right) \xrightarrow{D} N\left(0, \sigma^2 J_{\beta_0^{(r)}} B^- A B^- J_{\beta_0^{(r)}}^{\mathrm{T}}\right),$$

其中 B^- 表示矩阵 B 的广义逆,

$$A = E(VV^{\mathrm{T}}) - E[C(\beta_0^{\mathrm{T}}X)D^{-1}(\beta_0^{\mathrm{T}}X)C^{\mathrm{T}}(\beta_0^{\mathrm{T}}X)], \tag{7.2.12}$$

$$B = E(VV^{\mathrm{T}}) - E[C(\beta_0^{\mathrm{T}}X)\dot{\boldsymbol{g}}_0(\beta_0^{\mathrm{T}}X)E(X|\beta_0^{\mathrm{T}}X)^{\mathrm{T}}J_{\beta_0^{(r)}}],$$

$V = J_{\beta_0^{(r)}}^{\mathrm{T}} X \dot{\boldsymbol{g}}_0^{\mathrm{T}}(\beta_0^{\mathrm{T}}X) Z w(\beta_0^{\mathrm{T}}X)$, $C(u)$ 和 $D(u)$ 在条件 (C7.2.6) 中定义.

在模型 (7.1.1) 中, 如果 $q=1$ 且 $Z=1$, 那么 (7.1.1) 变为单指标模型. 由定理 7.2.2, 我们导出下列结果.

推论 7.2.1 在定理 7.2.2 的条件下, 如果在模型 (7.1.1) 中 $q=1$ 和 $Z=1$, 则

$$\sqrt{n}(\hat{\beta}-\beta_0) \xrightarrow{D} N(0, \sigma^2 J_{\beta_0^{(r)}} A_1^- J_{\beta_0^{(r)}}^{\mathrm{T}}),$$

其中 $A_1 = E\{J_{\beta_0^{(r)}}^{\mathrm{T}}[X - E(X|\beta_0^{\mathrm{T}}X)][X - E(X|\beta_0^{\mathrm{T}}X)]^{\mathrm{T}} J_{\beta_0^{(r)}} \dot{\boldsymbol{g}}_0^2(\beta_0^{\mathrm{T}}X) w(\beta_0^{\mathrm{T}}X)\}$.

注 7.2.1 对单指标模型, 用 $\hat{\beta}^*$ 表示由 Härdle(1993) 定义的 β_0 的估计量. 由 Härdle 等 (1993) 与 Xia 和 Li(1999) 给出的定理, 可知 $\hat{\beta}^*$ 的渐近协方差矩阵是广义逆 $\sigma^2 A_2^-$, 其中

$$A_2 = E\left\{[X - E(X|\beta_0^{\mathrm{T}}X)][X - E(X|\beta_0^{\mathrm{T}}X)]^{\mathrm{T}} \dot{\boldsymbol{g}}_0^2(\beta_0^{\mathrm{T}}X) w(\beta_0^{\mathrm{T}}X)\right\}.$$

可以证明, 存在 A_2 的下列形式的广义逆: $J_{\beta_0^{(r)}} A_1^- J_{\beta_0^{(r)}}^{\mathrm{T}}$, 使得 $A_2^- - J_{\beta_0^{(r)}} A_1^- J_{\beta_0^{(r)}}^{\mathrm{T}}$ 为一个半正定矩阵, 参看 Cui 等 (2011) 的推论 1. 这就证明了推论 7.2.1 中给出的协方差阵小于 $\sigma^2 A_2^-$. 因此, 推论 7.2.1 表明我们给出的估计量更有效. 参见定理 3.10.2 和定理 3.10.3.

因为 $\hat{\beta}$ 是 $\sqrt{n}$ 相合的, 立刻得到下列结果.

定理 7.2.3 设条件 (C7.2.1)~(C7.2.8) 成立, 则

$$\sqrt{nh}\left[\hat{\boldsymbol{g}}^*(u) - \boldsymbol{g}_0(u) - \frac{1}{2}h^2\mu_2\ddot{\boldsymbol{g}}_0(u)\right] \xrightarrow{D} N\left(0, f^{-1}(u)\nu_0\sigma^2 D^{-1}(u)\right),$$

其中 $\mu_2 = \int u^2 K(u)\mathrm{d}u$, $\nu_0 = \int K^2(u)\mathrm{d}u$.

对误差方差的估计量 $\hat{\sigma}^2$, 有下述结果.

定理 7.2.4 设条件 (C7.2.1)~(C7.2.8) 成立. 进一步, $K(u)$ 的傅里叶变换是绝对可积的, 且 X 的分布具有有界支撑, 则

$$\hat{\sigma}^2-\sigma^2=O\left(\left\{\frac{\log\log n}{n}\right\}^{1/2}\right),\quad \text{a.s.}$$

应用定理 7.2.2 构造 β_0 的置信域, 需要估计 A 和 B. 由 "嵌入" 方法, 矩阵 A 和 B 的相合估计可定义为

$$\hat{A}=\frac{1}{n}\sum_{i=1}^{n}\left[\hat{V}_i\hat{V}_i^{\mathrm{T}}-\hat{C}(\hat{\beta}^{\mathrm{T}}X_i)\hat{D}^{-1}(\hat{\beta}^{\mathrm{T}}X_i)\hat{C}^{\mathrm{T}}(\hat{\beta}^{\mathrm{T}}X_i)\right], \tag{7.2.13}$$

$$\hat{B}=\frac{1}{n}\sum_{i=1}^{n}\left[\hat{V}_i\hat{V}_i^{\mathrm{T}}-\hat{C}(\hat{\beta}^{\mathrm{T}}X_i)\hat{\boldsymbol{g}}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})\hat{\mu}^{\mathrm{T}}(\hat{\beta}^{\mathrm{T}}X_i)J_{\hat{\beta}^{(r)}}\right],$$

其中 $\hat{V}_i=J_{\hat{\beta}^{(r)}}^{\mathrm{T}}X_i\hat{\boldsymbol{g}}^{\mathrm{T}}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})Z_iw(\hat{\beta}^{\mathrm{T}}X_i)$, $\hat{C}(\cdot)=\sum_{i=1}^{n}W_{ni}(\cdot)\hat{V}_iZ_i^{\mathrm{T}}$, $\hat{D}(\cdot)=\sum_{i=1}^{n}W_{ni}(\cdot)Z_iZ_i^{\mathrm{T}}$, $\hat{\mu}(\cdot)=\sum_{i=1}^{n}W_{ni}(\cdot)X_i$. 这里

$$W_{ni}(\cdot)=K_1\left(\frac{\hat{\beta}^{\mathrm{T}}X_i-\cdot}{b_n}\right)\Big/\sum_{k=1}^{n}K_1\left(\frac{\hat{\beta}^{\mathrm{T}}X_k-\cdot}{b_n}\right),$$

$K_1(\cdot)$ 是核函数, b_n 是带宽且 $0<b_n\to 0$.

由定理 7.2.2 和定理 7.2.3, 可得

$$\left[\hat{\sigma}^2J_{\hat{\beta}^{(r)}}\hat{B}^{-1}\hat{A}\hat{B}^{-1}J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\right]^{-1/2}\sqrt{n}(\hat{\beta}-\beta_0)\xrightarrow{D}N(0,1).$$

使用 (Arnold, 1981) 的定理 10.2d 可得

$$(\hat{\beta}-\beta_0)^{\mathrm{T}}\left[n^{-1}\hat{\sigma}^2J_{\hat{\beta}^{(r)}}\hat{B}^{-1}\hat{A}\hat{B}^{-1}J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\right]^{-1}(\hat{\beta}-\beta_0)\xrightarrow{D}\chi_p^2. \tag{7.2.14}$$

(7.2.14) 可以用来构造 β_0 的置信域, 并可作假设检验.

7.3 经验似然

本节用经验似然方法研究模型 (7.1.1) 中参数 β_0 的置信域的构造问题. 构造 β_0 的估计的经验似然比和调整的经验似然比函数, 并证明它们的渐近性质.

7.3.1 估计的经验似然

为构造 β_0 的经验似然比函数, 引入辅助随机变量

$$\hat{Z}_i(\beta^{(r)}) = [Y_i - \hat{\boldsymbol{g}}^{\mathrm{T}}(\beta^{\mathrm{T}}X_i;\beta)Z_i]\hat{\dot{\boldsymbol{g}}}^{\mathrm{T}}(\beta^{\mathrm{T}}X_i;\beta)Z_iJ^{\mathrm{T}}_{\beta^{(r)}}X_iw(\beta^{\mathrm{T}}X_i), \tag{7.3.1}$$

其中 $\hat{\boldsymbol{g}}$ 和 $\hat{\dot{\boldsymbol{g}}}$ 由 (7.2.5) 式定义. 因此, 一个估计的经验对数似然比函数定义为

$$\hat{l}_{\mathrm{EEL}}(\beta^{(r)}) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\middle| p_i \geqslant 0, \sum_{i=1}^{n}p_i = 1, \sum_{i=1}^{n}p_i\hat{Z}_i(\beta^{(r)}) = 0\right\}.$$

下面的定理给出了 $\hat{l}_{\mathrm{EEL}}(\beta^{(r)})$ 的渐近性质.

定理 7.3.1 设条件 (C7.2.1)~(C7.2.8) 成立, 又设 $A_0 = E\{[J^{\mathrm{T}}_{\beta_0^{(r)}}X\dot{\boldsymbol{g}}_0^{\mathrm{T}}(\beta_0^{\mathrm{T}}X)Zw(\beta_0^{\mathrm{T}}X)]^{\otimes 2}\}$ 为正定矩阵, 则

$$\hat{l}_{\mathrm{EEL}}(\beta_0^{(r)}) \xrightarrow{D} w_1\chi^2_{1,1} + \cdots + w_{p-1}\chi^2_{1,p-1},$$

其中 $w_i(1 \leqslant i \leqslant p-1)$ 是矩阵 $M = A_0^{-1}A$ 的特征根, A 由 (7.2.12) 式定义, $\chi^2_{1,i}(1 \leqslant i \leqslant p-1)$ 是自由度为 1 的独立 χ^2_1 变量.

为应用定理 3.12.1 构造 $\beta_0^{(r)}$ 的置信域, 我们需要给出未知权 w_i 的相合估计, 即需要估计 A_0 和 A. 矩阵 A 的估计由 (7.2.13) 式定义, 亦可记作 $\hat{A}(\hat{\beta}^{(r)})$; A_0 的估计定义为 $\hat{A}_0(\hat{\beta}^{(r)}) = n^{-1}\sum_{i=1}^{n}(\hat{V}_i\hat{V}_i^{\mathrm{T}})$, 其中 $\hat{V}_i = J^{\mathrm{T}}_{\hat{\beta}^{(r)}}X_i\hat{\dot{\boldsymbol{g}}}^{\mathrm{T}}(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})Z_iw(\hat{\beta}^{\mathrm{T}}X_i)$, $\hat{\beta} = (\hat{\beta}_1,\cdots,\hat{\beta}_p)^{\mathrm{T}}$ 是由上节定义的 β_0 的估计方程估计, $\hat{\beta}^{(r)} = (\hat{\beta}_1,\cdots,\hat{\beta}_{r-1},\hat{\beta}_{r+1},\cdots,\hat{\beta}_p)^{\mathrm{T}}$, 这就推出 $\hat{M} = \hat{A}_0^{-1}\hat{A}$ 的特征根 $\hat{w}_i(1 \leqslant i \leqslant p-1)$ 是 w_i 的相合估计. 设 $H(\cdot)$ 为给定数据 $\{(Y_i, X_i, Z_i), 1 \leqslant i \leqslant n\}$ 下加权和 $\hat{s} = \hat{w}_1\chi^2_{1,1} + \cdots + \hat{w}_{p-1}\chi^2_{1,p-1}$ 的条件分布, 并设 $\hat{c}_\alpha$ 是 $H(\cdot)$ 的 $1-\alpha$ 分位数. 那么 $\beta_0^{(r)}$ 的近似 $1-\alpha$ 置信域为

$$\hat{I}_\alpha(\beta^{(r)}) = \{\beta^{(r)} | \hat{l}_{\mathrm{EEL}}(\beta^{(r)}) \leqslant \hat{c}_\alpha, ||\beta^{(r)}|| < 1\}.$$

在实际中, 利用 Monte-Carlo 模拟可以得到分布 $H(\cdot)$ 及其分位数 $\hat{c}_\alpha$.

7.3.2 调整的经验似然

本节给出一个调整的经验对数似然比, 其渐近分布是自由度为 $p-1$ 的标准 χ^2 分布. 为此, 定义调整因子

$$\hat{r}(\beta^{(r)}) = \frac{\mathrm{tr}\{\hat{A}^{-1}(\beta^{(r)})\hat{M}(\beta^{(r)})\}}{\mathrm{tr}\{\hat{A}_0^{-1}(\beta^{(r)})\hat{M}(\beta^{(r)})\}},$$

其中

$$\hat{M}(\beta^{(r)}) = \left[\sum_{i=1}^{n}\hat{Z}_i(\beta^{(r)})\right]\left[\sum_{i=1}^{n}\hat{Z}_i(\beta^{(r)})\right]^{\mathrm{T}}.$$

由此定义一个调整的经验对数似然比函数

$$\hat{l}_{\mathrm{AEL}}(\beta^{(r)}) = \hat{r}(\beta^{(r)})\hat{l}_{\mathrm{EEL}}(\beta^{(r)}). \tag{7.3.2}$$

下述定理给出了 $\hat{l}_{\mathrm{AEL}}(\beta^{(r)})$ 的渐近分布.

定理 7.3.2 在定理 7.3.1 的条件下, 有 $\hat{l}_{\mathrm{AEL}}(\beta_0^{(r)}) \xrightarrow{D} \chi^2_{p-1}$.

基于定理 7.3.2, 可以使用 $\hat{l}_{\mathrm{AEL}}(\beta^{(r)})$ 构造 $\beta_0^{(r)}$ 的置信域. 设

$$\hat{I}_{\mathrm{AEL},\alpha}(\beta^{(r)}) = \left\{\beta^{(r)} \middle| \hat{l}_{\mathrm{AEL}}(\beta^{(r)}) \leqslant \chi^2_{p-1}(1-\alpha), ||\beta^{(r)}|| < 1\right\},$$

那么 $\hat{I}_{\mathrm{AEL},\alpha}(\beta^{(r)})$ 给出 $\beta_0^{(r)}$ 的近似 $1-\alpha$ 置信域.

7.4 模拟和应用

本节首先进行模拟来展示上节的方法对有限数据集的执行情况. 然后将估计方法和单指标变系数模型应用到实际数据分析之中. 在下列例子中, 我们使用 Epanechnikov 核, 即 $K(x) = 0.75(1-x^2)I(|x| \leqslant 1)$.

7.4.1 模拟

例 7.4.1 考虑单指标变系数模型

$$Y_i = g_0(\beta_0^{\mathrm{T}}X_i) + g_1(\beta_0^{\mathrm{T}}X_i)Z_{i1} + g_2(\beta_0^{\mathrm{T}}X_i)Z_{i2} + \varepsilon_i, \tag{7.4.1}$$

其中 $\beta_0 = (\beta_{01}, \beta_{02}) = (1/2, \sqrt{3}/2)^{\mathrm{T}}$, $\{X_i = (X_{i1}, X_{i2})^{\mathrm{T}}, 1 \leqslant i \leqslant n\}$ 为服从 $[-1,1]^2$ 上的均匀分布的独立随机向量, $\{(Z_{i1}, Z_{i2})^{\mathrm{T}}, 1 \leqslant i \leqslant n\}$ 为二变量正态分布 $N(0,\Sigma)$, $\mathrm{var}(Z_{i1}) = \mathrm{var}(Z_{i2}) = 1$, $\mathrm{cov}(Z_{i1}, Z_{i2}) = 0.5$, $\{\varepsilon_i, 1 \leqslant i \leqslant n\}$ 为服从正态分布 $N(0, 0.6^2)$ 的独立随机变量, $g_0(u) = \exp(-u)$, $g_1(u) = 3u^3$, $g_2(u) = 5\cos(\pi u)$.

对 β_0 的估计, 我们使用两个方法作模拟: 估计方程 (EE) 和 Xia 和 Li(1999) 提出的最小二乘 (LS) 方法. 取权函数 $w(u) = I_{[-(1+\sqrt{3})/2, (1+\sqrt{3})/2]}(u)$. 分别在样本量 $n = 50, 100, 200$ 下通过 200 次运行计算偏差 (Bias)、标准差 (SD) 和根均方误差 (RMSE). 模拟结果放在表 7.4.1 中.

从表 7.4.1 可以看出, β_1 和 β_2 的基于 EE 的估计比基于 LS 的估计有更小的偏差和 RMSE. 这就表明 EE 方法改进了估计结果. 此外, 偏差、SD 和 RMSE 都随着样本量的增加而减少.

当样本量是 50 时, 200 个估计 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的直方图分别画在图 7.4.1(a) 和 (b) 中, 200 个估计 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的 Q–Q 图分别画在图 7.4.1(c) 和 (d) 中.

图 7.4.1 表明这些估计量是渐近正态的. $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 分别为 0.5001 和 0.8656, 它们的标准差分别为 0.0206 和 0.0118. σ^2 的估计为 0.3571.

表 7.4.1　例 7.4.1 的模拟结果. 估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 偏差、标准差和根均方误差

n	方法	$\hat{\beta}_1$			$\hat{\beta}_2$		
		Bias	SD	RMSE	Bias	SD	RMSE
50	EE	0.00156	0.02064	0.02064	−0.00042	0.01184	0.01185
	LS	0.00393	0.01955	0.01994	0.00596	0.01091	0.01243
100	EE	0.00055	0.01150	0.01151	−0.00042	0.00665	0.00666
	LS	0.00309	0.01162	0.01203	0.00677	0.00618	0.00916
200	EE	0.00088	0.00717	0.00723	−0.00055	0.00416	0.00419
	LS	0.00416	0.00697	0.00812	0.00644	0.00358	0.00737

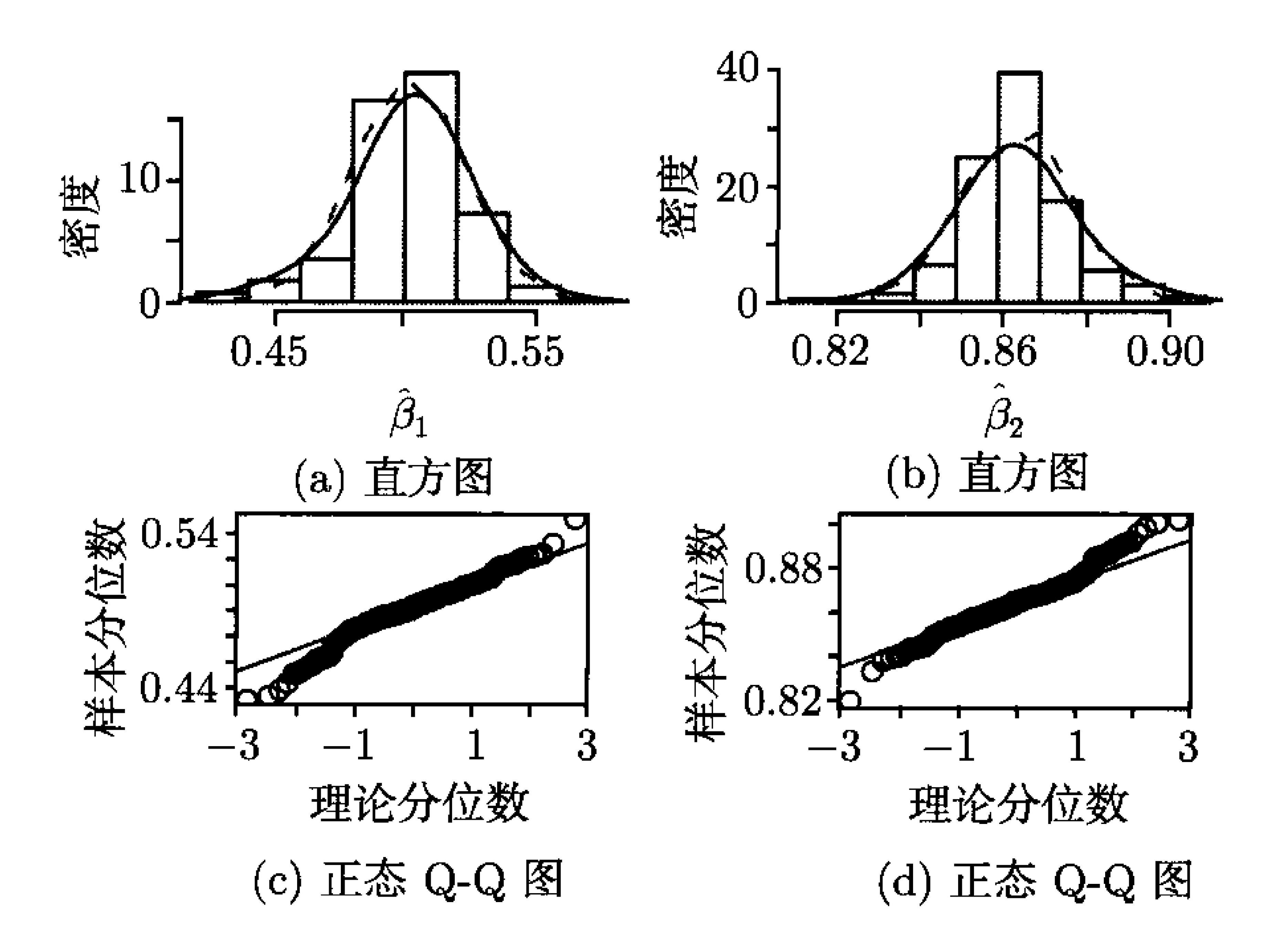


图 7.4.1　例 7.4.1 的模拟结果. (a) 对 $\hat{\beta}_1$ 和 (b) 对 $\hat{\beta}_2$: 每个参数的 200 个估计的直方图, 密度的估计曲线 (实线) 和正态密度曲线 (虚线); (c) 对 $\hat{\beta}_1$ 和 (d) 对 $\hat{\beta}_2$: 每个参数的 200 个估计的 Q-Q 图

我们也考虑了 $g_0(u)$, $g_1(u)$ 和 $g_2(u)$ 的估计. 图 7.4.2(a)~(c) 给出了真实的系数函数曲线, 估计的系数函数曲线和近似 0.95 逐点置信区间. 使用根均方误差对估计量 $\hat{g}_j^*(\cdot)$ 进行评估, 即 $\mathrm{RMSE}^2 = \mathrm{RMSE}_0^2 + \mathrm{RMSE}_1^2 + \mathrm{RMSE}_2^2$, 其中

$$\mathrm{RMSE}_j = \left[n_{\mathrm{grid}}^{-1} \sum_{k=1}^{n_{\mathrm{grid}}} \{\hat{g}_j^*(u_k) - g_j(u_k)\}^2 \right]^{1/2}, \quad j = 0, 1, 2,$$

$\{u_k, k = 1, \cdots, n_{\mathrm{grid}}\}$ 是适当的格子点. 当样本量为 50 时, 运行 200 次得到的 RMSE 的箱线图展示在图 7.4.2(d) 中.

从图 7.4.2(a)~(c) 可以看出, 每一个估计的曲线接近于真实系数函数曲线. 图 7.4.2(d) 表明未知函数的估计的 RMSE 都非常小.

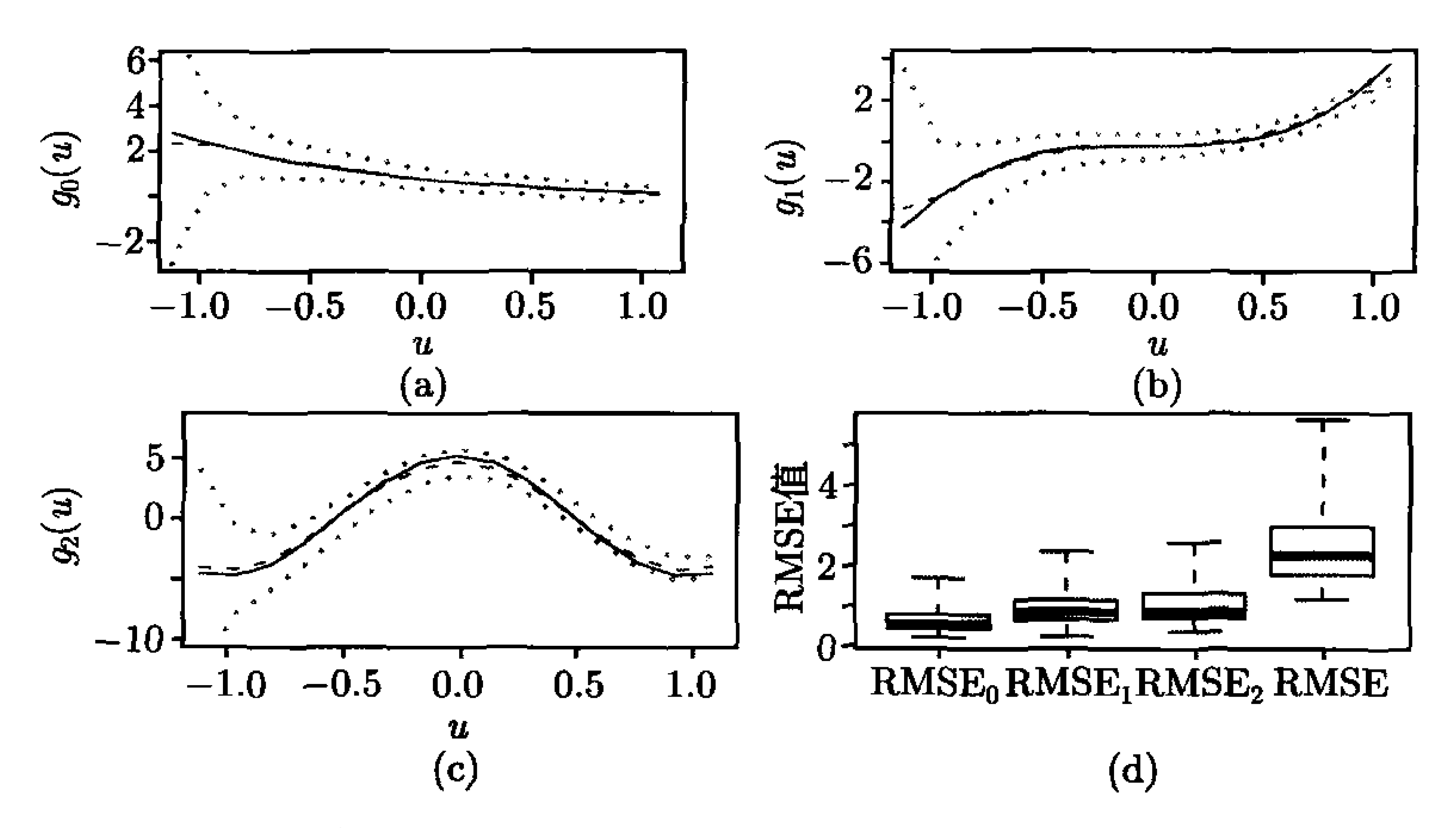


图 7.4.2 例 7.4.1 的模拟结果. 真实曲线 (实线), 估计的曲线 (虚线) 和近似 0.95 逐点置信区间. (a) 是对于 $g_0(\cdot)$, (b) 是对于 $g_1(\cdot)$, (c) 是对于 $g_2(\cdot)$; (d) 200 个 $\hat{g}_0(\cdot)$, $\hat{g}_1(\cdot)$ 和 $\hat{g}_2(\cdot)$ 的 RMSE 的箱线图

例 7.4.2 考虑单指标变系数模型

$$\begin{aligned}Y_t =& \exp(\{-8(0.8Y_{t-1}+0.6Y_{t-2}-0.6)^2\}\\&+0.5\sin(2\pi(0.8Y_{t-1}+0.6Y_{t-2}-0.6))Y_{t-1}+\varepsilon_t,\end{aligned}$$

其中 $Z_t=(1,Y_{t-1})^{\mathrm{T}}$, $X_t=(Y_{t-1},Y_{t-2})^{\mathrm{T}}$, $t>2$, $\beta_0=(0.8,0.6)^{\mathrm{T}}$, 诸 ε_t 服从正态分布 $N(0,0.1^2)$.

取权函数 $w(u)=I_{[-4,4]}(u)$. 样本量为 50, 进行 200 次独立计算, 其结果列在表 7.4.2 中.

表 7.4.2 例 7.4.2 的模拟结果. 估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 偏差、标准差和根均方误差

n	方法	$\hat{\beta}_1$			$\hat{\beta}_2$		
		Bias	SD	RMSE	Bias	SD	RMSE
50	EE	−0.0608	0.0621	0.0869	0.0676	0.0642	0.0932
	LS	−0.0792	0.0712	0.1064	0.0754	0.0706	0.1033
100	EE	−0.0461	0.0541	0.0711	0.0522	0.0581	0.0781
	LS	−0.0502	0.0601	0.0783	0.0692	0.0617	0.0927
200	EE	−0.0332	0.0422	0.0537	0.0389	0.0460	0.0603
	LS	−0.0512	0.0578	0.0772	0.0574	0.0547	0.0793

从表 7.4.2 可以看出, 对于这个模型, β_0 基于 EE 的估计比基于 LS 的估计有更小的偏差、SD 和 RMSE.

例 7.4.3 考虑下列模拟的模型:

$$Y_i=3\exp\{-(\beta_0^{\mathrm{T}}X_i)^2\}+0.8(\beta_0^{\mathrm{T}}X_i)Z_{i1}+1.5\sin(2\pi\beta_0^{\mathrm{T}}X_i)Z_{i2}+\sigma\varepsilon_i,$$

其中 $X_i = (X_{i1}, X_{i2}, X_{i3})^{\mathrm{T}}$, $i = 1, \cdots, n$ 为服从 $[-1,1]^3$ 上均匀分布的独立随机向量, $(Z_{i1}, Z_{i2})^{\mathrm{T}} = (X_{i2}, X_{i3})^{\mathrm{T}}$, $\beta_0 = (1/3, 2/3, 2/3)^{\mathrm{T}}$, 诸 ε_i 是独立的标准正态变量.

使用内积 $|\hat{\beta}^{\mathrm{T}}\beta_0|$ 度量 $\hat{\beta}$ 的估计的精度. 取权函数 $w(u) = I_{[-5/3,5/3]}(u)$. 对样本量 $n = 50, 100, 200$ 和噪声大小 $\sigma = 0.5, 1, 1.5$ 的每一种组合作 200 次运行, 其结果列在表 7.4.3 中; 并且绝对内积的三种箱线图也展示在图 7.4.3 中.

表 7.4.3 例 7.4.3 的模拟结果. 估计量 $\hat{\beta}$ 和真值 β_0 之间的内积的均值和标准差

n	方法	$\sigma_\varepsilon = 0.5$ Mean(SD)	$\sigma_\varepsilon = 1$ Mean(SD)	$\sigma_\varepsilon = 1.5$ Mean(SD)
50	EE	0.9756(0.0245)	0.9453(0.0403)	0.9377(0.0510)
	LS	0.8661(0.0410)	0.8508(0.0660)	0.8345(0.1013)
100	EE	0.9898(0.0113)	0.9543(0.0384)	0.9441(0.0428)
	LS	0.8714(0.0287)	0.8522(0.0471)	0.8399(0.0845)
200	EE	0.9955(0.0069)	0.9713(0.0199)	0.9464(0.0407)
	LS	0.8737(0.0158)	0.8709(0.0394)	0.8443(0.0759)

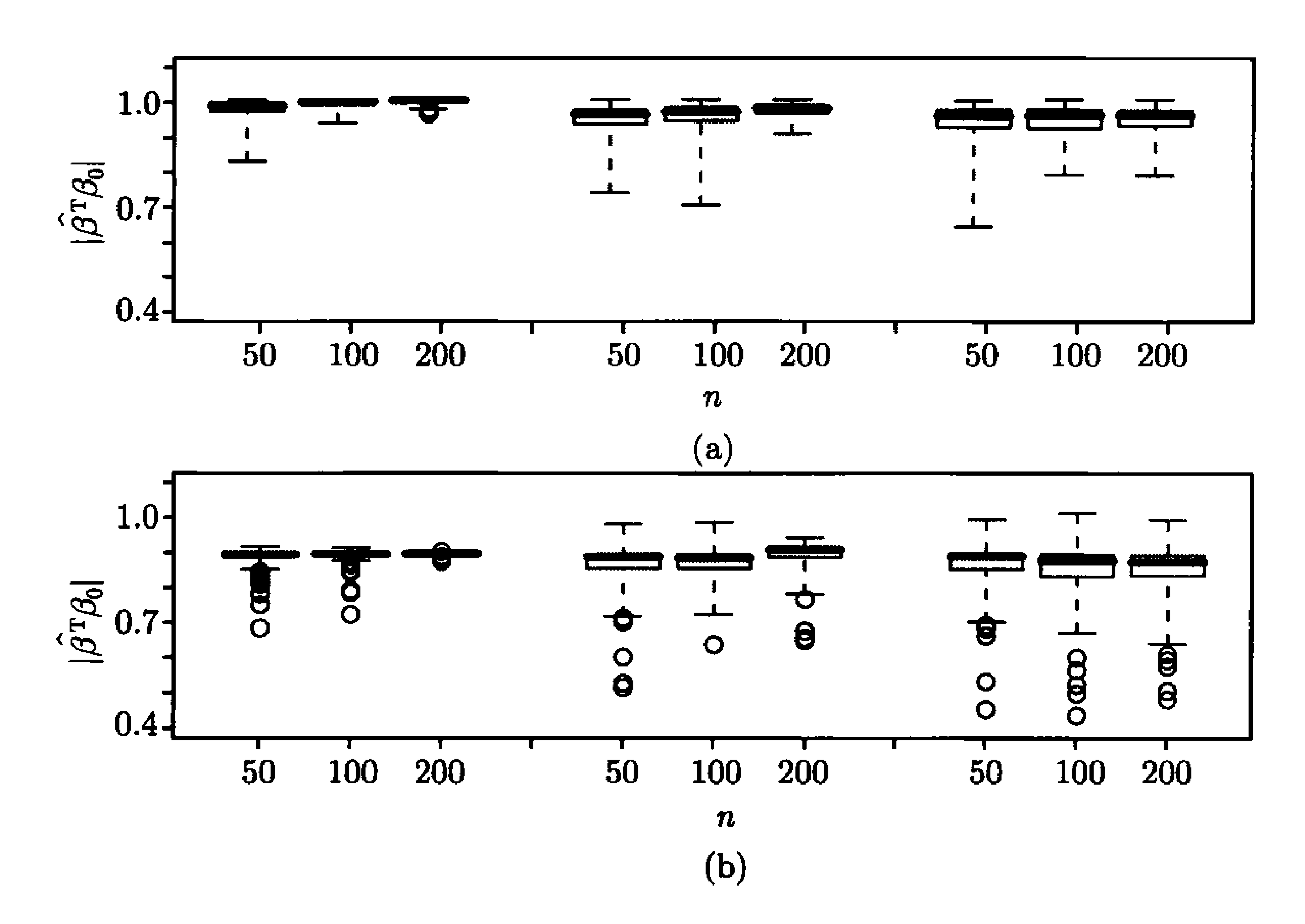


图 7.4.3 例 7.4.3 的模拟结果. 对 $\sigma = 0.5, 1, 1.5$ 和样本量 $n = 50, 100, 200$, 绝对内积 $|\hat{\beta}^{\mathrm{T}}\beta_0|$ 的 200 次运行的三组箱线图. (a) 是对于 EE, (b) 是对于 Xia 等 (1999) 提出的 LS

从图 7.4.3 可以看出, 基于 EE 的估计的内积的均值比基于 LS 的内积的均值大, 并且他们的标准差 (在括弧内) 比基于 LS 的更小. 与 Xia 等 (1999) 进行比较, 图 7.4.3 中值的分布比他们的更接近于 1.

我们使用三种方法构造 β_0 的置信域: 估计的经验似然 (EEL), 调整的经验似

然 (AEL) 和正态逼近 (NA). 通过 200 次运行计算 β_0 的近似为 0.95 置信域及其覆盖概率, 由此比较三种方法的优劣. 模拟结果展示在图 7.4.4 和表 7.4.4 中.

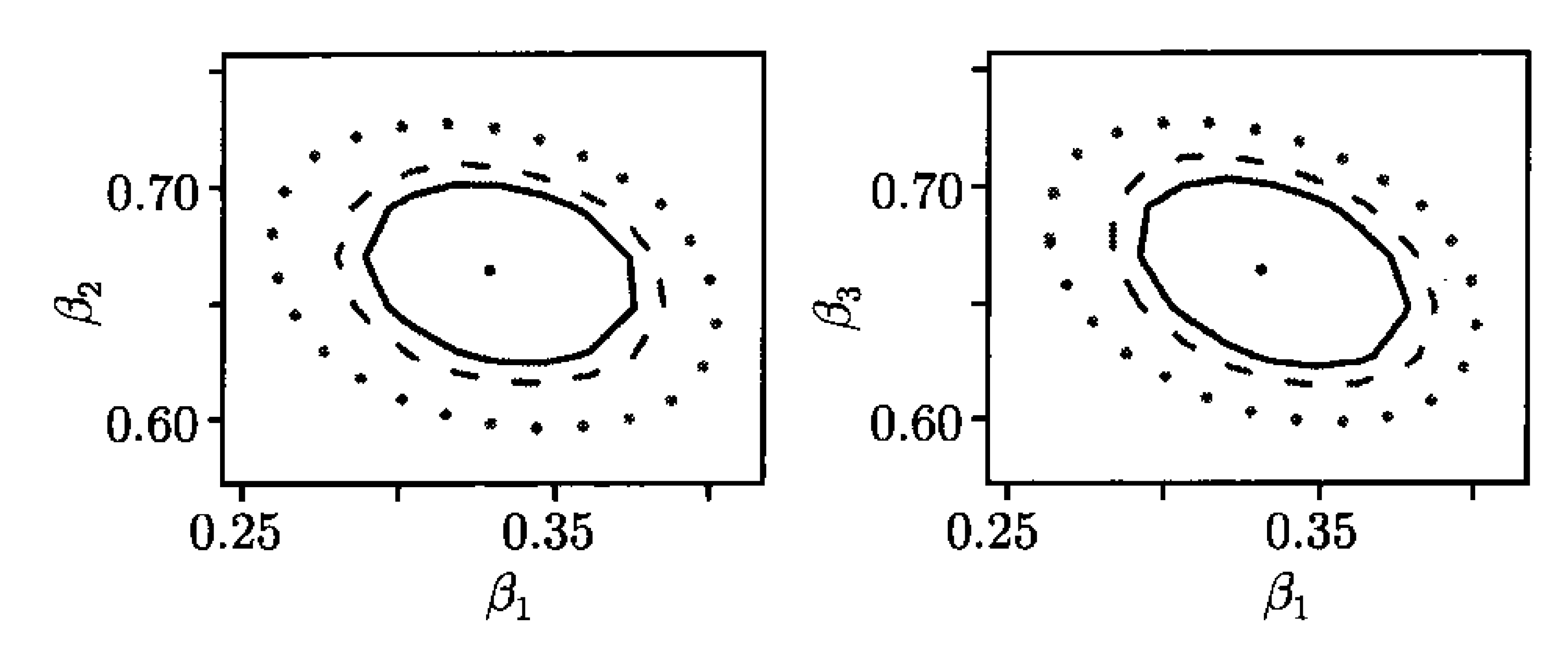


图 7.4.4 基于 AEL(实线)、EEL(虚线) 和 NA(点线) 及样本量 $n = 100$, (β_1, β_2) 和 (β_1, β_3) 的近似 0.95 置信域

表 7.4.4 参数 (β_1, β_2) 和 (β_1, β_3) 的近似 0.95 置信域的覆盖概率

n	(β_1, β_2)			(β_1, β_3)		
	AEL	EEL	NA	AEL	EEL	NA
100	0.9067	0.9053	0.8937	0.9067	0.9053	0.8970
200	0.9257	0.9247	0.9190	0.9257	0.9247	0.9177
300	0.9510	0.9417	0.9327	0.9510	0.9417	0.9340

从图 7.4.4 看出, AEL 的效果最好, 它的置信域比其他两个方法的置信域小. 两种经验似然置信域的覆盖概率有相当小的差别, 但他们都比 NA 的实施效果好.

7.4.2 实例分析

例 7.4.4 环境数据. 我们使用的数据集包含意大利的米兰市的死亡人数和空气污染的日常测量. 两个污染源 —— 二氧化硫 (SO_2) 和周围空气中的悬浮粒子 (TSP). 该数据集包含连续 1434 天的观察结果 (自 1980 年 1 月 1 日起).

Ruppert 等 (2003) 发现上述环境数据集遵循一个非线性时间序列. 我们的主要兴趣是研究化学污染程度和每天呼吸病死亡人数 (RD) 的相关性. 设 $Y_i = \sqrt{\text{RD}}/10$, $X_{i1} = \sqrt{\text{SO}_2}/10$, $X_{i2} = \sqrt{\text{TSP}}/10$. 我们用单指标变系数模型 (SIVCM)(7.1.1) 来拟合他们的相关性. 其中 $(Z_{i1}, Z_{i2})^{\mathrm{T}} = (X_{i1}, X_{i2})^{\mathrm{T}}$. 为了更好地比较, 我们也使用下列单指标模型 (SIM) 和线性模型 (LM):

$$Y_i = g(\beta_1 X_{i1} + \beta_2 X_{i2}) + \varepsilon_i, \tag{7.4.2}$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i.$$

当使用 LM 拟合这个数据集时, 我们发现 β_2 的估计是 0.001584, 对假设 $H_0 : \beta_2 = 0$, $H_1 : \beta_2 \neq 0$, 具有大于 0.05 的 p 值 0.836, 线性相关性是不显著的. 通过分析得到

的 R^2 是 0.2144, 是不显著相关的. 因此, 我们使用 SIVCM 和 SIM 分析这个数据集. 取权函数 $w(u) = 1$. 计算结果展示在表 7.4.5 和图 7.4.5 中.

表 7.4.5　例 7.4.4 的模拟结果. β_0, β_1, β_2 和 σ^2 的估计及最优带宽 $h_{\rm opt}$

模型	$\hat\beta_0$	$\hat\beta_1$	$\hat\beta_2$	$\hat\sigma^2$	$h_{\rm opt}$
SIVCM		0.9746	−0.2241	0.0035	0.2438
SIM		0.9724	0.2332	0.0036	0.1531
LM	0.5161	0.0552	0.0016	0.0038	

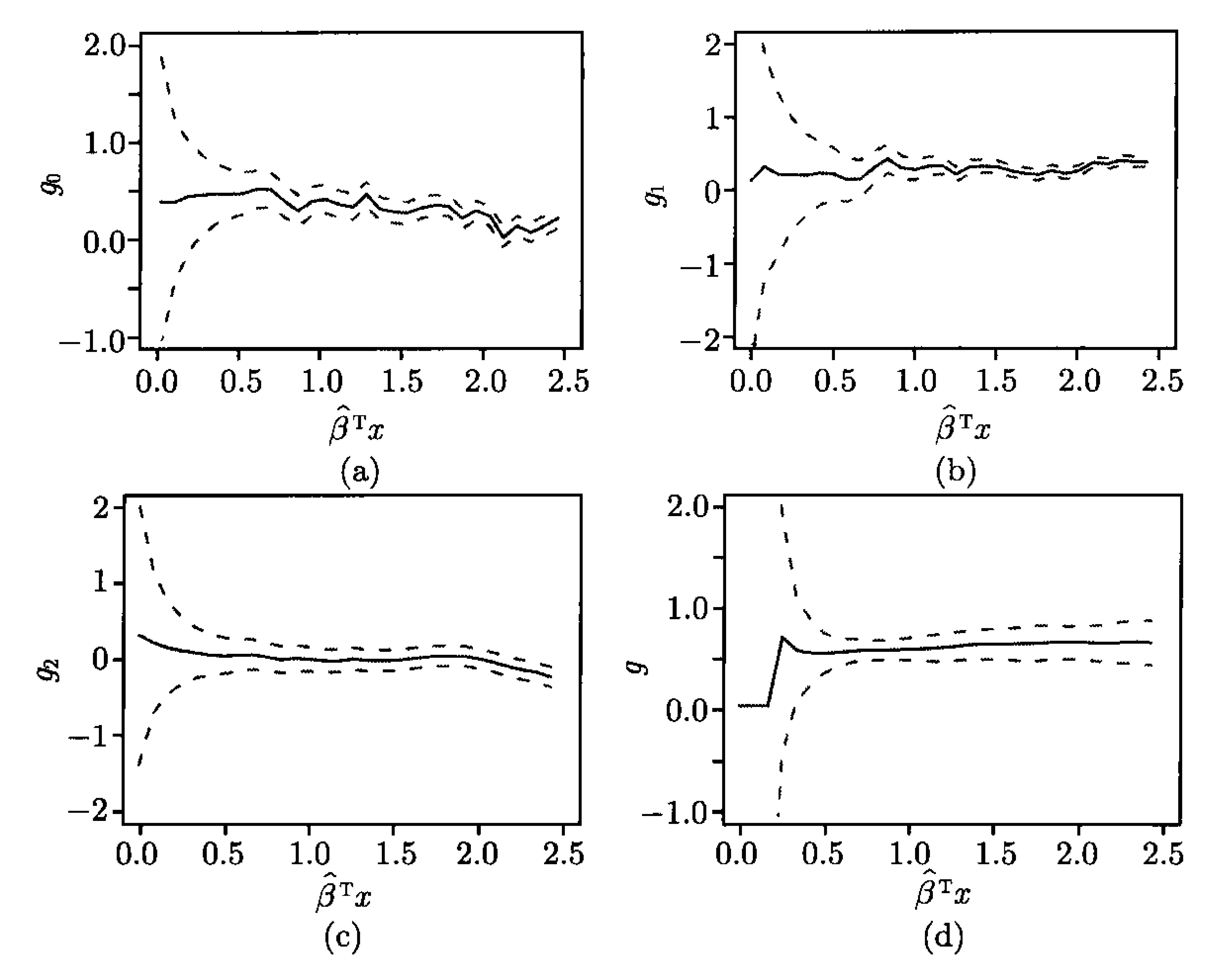


图 7.4.5　例 7.4.4 的模拟结果. (a)~(d) 内的实曲线为 $g_0(\cdot)$, $g_1(\cdot)$ 和 $g_2(\cdot)$ 的估计的曲线; 虚线表示近似 0.95 逐点置信区间. (a)~(c) 对于模型 (7.4.1), (d) 对于模型 (7.4.2)

从表 7.4.5 可以看出, SIVCM 比 SIM 和 LM 给出了更小的拟合误差的方差. 图 7.4.5 给出了估计的系数函数和联系函数, 并表明用常系数回归模型拟合这个数据集的建议是不合适的. 此外, 从图 7.4.5(d) 看出, 每天呼吸病死亡人数随着空气污染指标 $\hat\beta^{\rm T}X$ 的增加而增加, 其含义是增加空气污染将会增加呼吸疾病的发作.

7.5　定理的证明

下文用 c 表示任何正的常数, 每次出现时可以取不同的值. 下列引理类似于 (Xia and Li, 1999) 的引理 2, 这里略去它的证明.

引理 7.5.1 设函数 $\varphi(u)$ 在闭区间 $[a,b]$ 上具有有界导数, $\{(U_i,\xi_i)\}$ 是严平稳强混合序列且具有混合系数 $\alpha(n)=O(\rho^n)$, $0<\rho<1$ 为某个常数. 对任何 $r>1$, $E|U_i|^r<\infty$, $E|\xi_i|^r<\infty$, 对任何 $l>1$, 条件密度 $f_{U_1,U_l|\xi_1,\xi_l}(u_1,u_l|v_1,v_l)$ 和 $f_{U_1|\xi_1}(u|v)$ 有界. 如果条件 (C7.2.4) 成立, 则

$$\sup_{u\in[a,b]}\left|\frac{1}{n}\sum_{i=1}^{n}[\psi_i(u)-E\{\psi_i(u)\}]\right|=O\left(\left\{\frac{\log n}{nh}\right\}^{1/2}\right),\quad \text{a.s},$$

其中 $\psi_i(u)=K_h(U_i-u)\varphi(U_i)\xi_i$.

定理 7.2.1 的证 由 (7.2.6) 式和引理 7.5.1, 我们可以证明: 对 $u\in\mathcal{U}_w$ 和 $\beta\in\mathcal{B}_n$, 一致成立

$$S_{n,j}(u;\beta)=f(u)D(u)\mu_j+O\left(\left\{\frac{\log n}{nh}\right\}^{1/2}+h\right),\quad \text{a.s},\tag{7.5.1}$$

其中 $j=0,1,2,3$, $D(u)$ 在条件 (C7.2.6) 中定义. 由此立刻得到: 对 $u\in\mathcal{U}_w$ 和 $\beta\in\mathcal{B}_n$, 一致成立

$$S_n(u;\beta)=S(u)+O\left(\left\{\frac{\log n}{nh}\right\}^{1/2}+h\right),\quad \text{a.s},$$

其中 $S(u)=f(u)D(u)\otimes\mathrm{diag}\{1,\mu_2\}$, $\otimes$ 是 Kronecker 积. 使用事实

$$(H_1+hH_2)^{-1}=H_1^{-1}-hH_1^{-1}H_2H_1^{-1}+o(h^2),\tag{7.5.2}$$

可得, 对 $u\in\mathcal{U}_w$ 和 $\beta\in\mathcal{B}_n$ 一致成立

$$S_n^{-1}(u;\beta)=S^{-1}(u)+O\left(\left\{\frac{\log n}{nh}\right\}^{1/2}+h\right),\quad \text{a.s.}\tag{7.5.3}$$

记

$$\xi_{n,j}^*(u;\beta)=\frac{1}{n}\sum_{i=1}^{n}Z_i\{Y_i-\boldsymbol{g}_0^{\mathrm{T}}(\beta^{\mathrm{T}}X_i)Z_i\}\left(\frac{\beta^{\mathrm{T}}X_i-u}{h}\right)^j K_h(\beta^{\mathrm{T}}X_i-u),$$

$$\xi_n^*\equiv\xi_n^*(u;\beta)=\begin{pmatrix}\xi_{n,0}^*(u;\beta)\\ \xi_{n,1}^*(u;\beta)\end{pmatrix}.$$

由引理 7.5.1 并类似于上面的证明可以得到: 对 $u\in\mathcal{U}_w$ 和 $\beta\in\mathcal{B}_n$ 一致成立

$$\xi_n^*(u;\beta)=O\left(\left\{\frac{\log n}{nh}\right\}^{1/2}+n^{-1/2}\right),\quad \text{a.s.}\tag{7.5.4}$$

由 (7.2.7) 式并对 $\boldsymbol{g}_0(\beta^{\mathrm{T}}X_i)$ 在 u 处使用 Taylor 公式可得: 对 $u\in\mathcal{U}_w$ 和 $\beta\in\mathcal{B}_n$ 一致成立

$$\xi_{n,0}-\xi_{n,0}^*=S_{n,0}\boldsymbol{g}_0(u)+S_{n,1}h\dot{\boldsymbol{g}}_0(u)+\frac{1}{2}h^2S_{n,2}\ddot{\boldsymbol{g}}_0(u)+o(h^2),\quad \text{a.s.}$$

$$\xi_{n,1}-\xi_{n,1}^*=S_{n,1}\boldsymbol{g}_0(u)+S_{n,2}h\dot{\boldsymbol{g}}_0(u)+\frac{1}{2}h^2S_{n,3}\ddot{\boldsymbol{g}}_0(u)+o(h^2).\quad \text{a.s.}$$

因此, 对 $u\in\mathcal{U}_w$ 和 $\beta\in\mathcal{B}_n$ 一致成立

$$\xi_n-\xi_n^*=S_n\begin{pmatrix}\boldsymbol{g}_0(u)\\ h\dot{\boldsymbol{g}}_0(u)\end{pmatrix}+\frac{1}{2}h^2\begin{pmatrix}S_{n,2}\ddot{\boldsymbol{g}}_0(u)\\ S_{n,3}\ddot{\boldsymbol{g}}_0(u)\end{pmatrix}+o(h^2),\quad \text{a.s.}\qquad (7.5.5)$$

其中 $S_n=S_n(u;\beta)$ 和 $S_{n,j}=S_{n,j}(u;\beta)$, $j=2,3$. 于是, 由 (7.5.1) 和 (7.5.3)~(7.5.5) 可以推出: 对 $u\in\mathcal{U}_w$ 和 $\beta\in\mathcal{B}_n$ 一致成立

$$\begin{pmatrix}\hat{\boldsymbol{g}}(u;\beta)-\boldsymbol{g}_0(u)\\ h\{\hat{\dot{\boldsymbol{g}}}(u;\beta)-\dot{\boldsymbol{g}}_0(u)\}\end{pmatrix}=S^{-1}\xi_n^*+\frac{1}{2}h^2\begin{pmatrix}\mu_2\ddot{\boldsymbol{g}}_0(u)\\ \dfrac{\mu_3}{\mu_2}\ddot{\boldsymbol{g}}_0(u)\end{pmatrix}+o(c_n),\quad \text{a.s.}\qquad (7.5.6)$$

其中 $S=S(u)$, $c_n=n^{-1/2}+h^2$. 最后结合 (7.5.4) 和 (7.5.6) 式即可完成定理 7.2.1 的证明.

对某个正的常数 c_2, 设 $\mathcal{B}_n^{(r)}=\{\beta^{(r)}:\|\beta^{(r)}-\beta_0^{(r)}\|\leqslant c_2n^{-1/2}\}$. 记 $\mathcal{G}=\{\boldsymbol{g}:\mathcal{U}_w\times\mathcal{B}\mapsto\mathbf{R}^q\}$, $\|\boldsymbol{g}\|_{\mathcal{G}}=\sup_{u\in\mathcal{U}_w,\beta^{(r)}\in\mathcal{B}_n^{(r)}}\|\boldsymbol{g}(u;\beta(\beta^{(r)})\|$. 因为当 $\|\beta\|=1$ 时, $\|\beta-\beta_0\|\leqslant c_1n^{-1/2}$ 和 $\|\beta^{(r)}-\beta_0^{(r)}\|\leqslant c_2n^{-1/2}$ 是等价的, 于是由定理 7.2.1 可得 $\|\hat{\boldsymbol{g}}-\boldsymbol{g}_0\|_{\mathcal{G}}=o_P(1)$, $\|\hat{\dot{\boldsymbol{g}}}-\dot{\boldsymbol{g}}_0\|_{\mathcal{G}}=o_P(1)$, 因此可以假定: 对 $\delta=\delta_n\to0$ 和 $\delta>0$, $\boldsymbol{g}$ 位于 $\mathcal{G}_\delta$ 内, 其中

$$\mathcal{G}_\delta=\{\boldsymbol{g}\in\mathcal{G}:\|\boldsymbol{g}-\boldsymbol{g}_0\|_{\mathcal{G}}\leqslant\delta,\|\dot{\boldsymbol{g}}-\dot{\boldsymbol{g}}_0\|_{\mathcal{G}}\leqslant\delta\}.\qquad (7.5.7)$$

设 $\boldsymbol{g}_0(\beta^{\mathrm{T}}X;\beta)=E[\boldsymbol{g}_0(\beta_0^{\mathrm{T}}X)|\beta^{\mathrm{T}}X]$, $\dot{\boldsymbol{g}}_0(\beta^{\mathrm{T}}X;\beta)=E[\dot{\boldsymbol{g}}_0(\beta_0^{\mathrm{T}}X)|\beta^{\mathrm{T}}X]$,

$$Q(\boldsymbol{g},\beta^{(r)})=E\{[Y-\boldsymbol{g}^{\mathrm{T}}(\beta^{\mathrm{T}}X;\beta)Z]\dot{\boldsymbol{g}}^{\mathrm{T}}(\beta^{\mathrm{T}}X;\beta)ZJ_{\beta^{(r)}}^{\mathrm{T}}Xw(\beta^{\mathrm{T}}X)\},$$

$$Q_n(\boldsymbol{g},\beta^{(r)})=\frac{1}{n}\sum_{i=1}^n[Y_i-\boldsymbol{g}^{\mathrm{T}}(\beta^{\mathrm{T}}X_i;\beta)Z_i]\dot{\boldsymbol{g}}^{\mathrm{T}}(\beta^{\mathrm{T}}X_i;\beta)Z_iJ_{\beta^{(r)}}^{\mathrm{T}}X_iw(\beta^{\mathrm{T}}X_i).\qquad (7.5.8)$$

易知 $\boldsymbol{g}_0(\beta_0^{\mathrm{T}}X;\beta_0)=\boldsymbol{g}_0(\beta_0^{\mathrm{T}}X)$, $\dot{\boldsymbol{g}}_0(\beta_0^{\mathrm{T}}X;\beta_0)=\dot{\boldsymbol{g}}_0(\beta_0^{\mathrm{T}}X)$. 定义 $Q(g,\beta^{(r)})$ 在 $g_0(\cdot;\beta)$ 处关于 $\boldsymbol{g}(\cdot;\beta)$ 的函数导数:

$$\begin{aligned}&\varpi(\boldsymbol{g}_0(\cdot;\beta),\beta^{(r)})[\boldsymbol{g}(\cdot;\beta)-\boldsymbol{g}_0(\cdot;\beta)]\\&=\lim_{\tau\to0}[Q(\boldsymbol{g}_0(\cdot;\beta)+\tau(\boldsymbol{g}(\cdot;\beta)-\boldsymbol{g}_0(\cdot;\beta)),\beta^{(r)})-Q(\boldsymbol{g}_0(\cdot;\beta),\beta^{(r)})]\cdot\frac{1}{\tau}.\end{aligned}\qquad (7.5.9)$$

可得

$$\begin{aligned}&\varpi(\boldsymbol{g}_0(\beta^{\mathrm{T}}X;\beta),\beta^{(r)})(\boldsymbol{g}(\beta^{\mathrm{T}}X;\beta)-\boldsymbol{g}_0(\beta^{\mathrm{T}}X;\beta))\\=&-E\{[\boldsymbol{g}(\beta^{\mathrm{T}}X;\beta)-\boldsymbol{g}_0(\beta^{\mathrm{T}}X;\beta)]^{\mathrm{T}}Z\dot{\boldsymbol{g}}_0^{\mathrm{T}}(\beta^{\mathrm{T}}X;\beta)ZJ_{\beta^{(r)}}^{\mathrm{T}}Xw(\beta^{\mathrm{T}}X)\}.\end{aligned}\tag{7.5.10}$$

引理 7.5.2 设条件 (C7.2.1)~(C7.2.6) 成立, 则

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^{(r)}}||M_1(\beta^{(r)})||=o_P(n^{-1/2}),\tag{7.5.11}$$

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^{(r)}}||M_2(\beta^{(r)})||=o_P(n^{-1/2}),\tag{7.5.12}$$

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^{(r)}}||M_3(\beta^{(r)})||=o(n^{-1/2}),\tag{7.5.13}$$

$$\sqrt{n}M_4(\beta_0^{(r)})\xrightarrow{D}N(0,\sigma^2A),\tag{7.5.14}$$

其中 A 在定理 7.2.2 中定义

$$M_1(\beta^{(r)})=Q_n(\hat{\boldsymbol{g}},\beta^{(r)})-Q(\hat{\boldsymbol{g}},\beta^{(r)})-Q_n(\boldsymbol{g}_0,\beta_0^{(r)}),$$

$$\begin{aligned}M_2(\beta^{(r)})=&Q(\hat{\boldsymbol{g}},\beta^{(r)})-Q(\boldsymbol{g}_0,\beta^{(r)})\\&-\varpi(\boldsymbol{g}_0(\beta^{\mathrm{T}}X;\beta),\beta^{(r)})[\hat{\boldsymbol{g}}(\beta^{\mathrm{T}}X;\beta)-\boldsymbol{g}_0(\beta^{\mathrm{T}}X;\beta)],\end{aligned}$$

$$\begin{aligned}M_3(\beta^{(r)})=&\varpi(\boldsymbol{g}_0(\beta^{\mathrm{T}}X;\beta),\beta^{(r)})[\hat{\boldsymbol{g}}(\beta^{\mathrm{T}}X;\beta)-\boldsymbol{g}_0(\beta^{\mathrm{T}}X;\beta)]\\&-\varpi(\boldsymbol{g}_0(\beta_0^{\mathrm{T}}X),\beta_0^{(r)})[\hat{\boldsymbol{g}}(\beta_0^{\mathrm{T}}X;\beta_0)-\boldsymbol{g}_0(\beta_0^{\mathrm{T}}X)],\end{aligned}$$

$$M_4(\beta_0^{(r)})=Q_n(\boldsymbol{g}_0,\beta_0^{(r)})+\varpi(\boldsymbol{g}_0(\beta_0^{\mathrm{T}}X),\beta_0^{(r)})[\hat{\boldsymbol{g}}((\beta_0^{\mathrm{T}}X;\beta_0)-\boldsymbol{g}_0(\beta_0^{\mathrm{T}}X)].$$

证 首先证明 (7.5.11). 记 $r_n(\boldsymbol{g},\beta^{(r)})=\sqrt{n}[Q_n(\boldsymbol{g},\beta^{(r)})-Q(\boldsymbol{g},\beta^{(r)})]$. 注意到 $Q(\boldsymbol{g}_0,\beta_0^{(r)})=0$, 显然有

$$M_1(\beta^{(r)})=n^{-1/2}[r_n(\hat{\boldsymbol{g}},\beta^{(r)})-r_n(\boldsymbol{g}_0,\beta_0^{(r)})].\tag{7.5.15}$$

可以证明经验过程 $\{r_n(\boldsymbol{g},\beta^{(r)}):\boldsymbol{g}\in\mathcal{G}_1,\beta^{(r)}\in\mathcal{B}_1^{(r)}\}$ 具有随机等度连续性, 其中 $\mathcal{B}_1^{(r)}=[\beta^{(r)}:||\beta^{(r)}-\beta_0^{(r)}||\leqslant c_2]$ 和 $\mathcal{G}_1$ 在 (7.5.7) 式中定义, 其中 $\delta=1$, 它们分别是 $\mathcal{B}$ 和 $\mathcal{G}$ 的子集, 且满足 (7.5.11) 的证明的需要, 因为对充分大的 n, 由 $\delta\to 0$ 知 $\delta<1$. 通过验证 (Doukhan et al., 1995) 中定理 1 的条件可以推出上述随机等度连续性. 因此, 对 $\boldsymbol{g}\in\mathcal{G}_1$ 和 $\beta^{(r)}\in\mathcal{B}_1^{(r)}$, 一致成立 $r_n(\boldsymbol{g},\beta^{(r)})-r_n(\boldsymbol{g}_0,\beta_0^{(r)})=o_P(1)$. 结合 (7.5.15) 式和定理 7.2.2 即可证得 (7.5.11) 式.

下面证明 (7.5.12) 式. 由 (7.5.10) 式可以推出

$$M_2(\beta^{(r)}) = -E\Big[[\hat{\boldsymbol{g}}(\beta^{\mathrm{T}}X;\beta) - \boldsymbol{g}_0(\beta_0^{\mathrm{T}}X)]^{\mathrm{T}}Z \times [\hat{\dot{\boldsymbol{g}}}(\beta^{\mathrm{T}}X;\beta) - \dot{\boldsymbol{g}}_0(\beta^{\mathrm{T}}X;\beta)]^{\mathrm{T}}ZJ_{\beta^{(r)}}^{\mathrm{T}}Xw(\beta^{\mathrm{T}}X)\Big],$$

因此, 由定理 7.2.1 和条件 (C7.2.3) 不难得到: 对 $\beta^{(r)} \in \mathcal{B}_n^{(r)}$ 一致成立 $M_2(\beta^{(r)}) = o_P(n^{-1/2})$. 这就证明了 (7.5.12) 式.

下面证明 (7.5.13) 式. 记 $\varphi(\beta^{(r)}) = [\hat{\boldsymbol{g}}(\beta^{\mathrm{T}}X;\beta) - \boldsymbol{g}_0(\beta^{\mathrm{T}}X;\beta)]^{\mathrm{T}}Z$, $\psi(\beta^{(r)}) = \dot{\boldsymbol{g}}_0^{\mathrm{T}}(\beta^{\mathrm{T}}X;\beta)ZJ_{\beta^{(r)}}^{\mathrm{T}}Xw(\beta^{\mathrm{T}}X)$. 由 (7.5.10) 式推得

$$\begin{aligned}M_3(\beta^{(r)}) &= -E[\varphi(\beta^{(r)})\psi(\beta^{(r)})] + E[\varphi(\beta_0^{(r)})\psi(\beta_0^{(r)})]\\ &= -E\{[\varphi(\beta^{(r)}) - \varphi(\beta_0^{(r)})]\psi(\beta^{(r)})\}\\ &\quad - E\{\varphi(\beta_0^{(r)})[\psi(\beta^{(r)}) - \psi(\beta_0^{(r)})]\}\\ &\equiv M_{31}(\beta^{(r)}) + M_{32}(\beta^{(r)}).\end{aligned}$$

由条件 (C7.2.2) 得到

$$\begin{aligned}&||\varphi(\beta^{(r)}) - \varphi(\beta_0^{(r)})||\\ =&||\left\{[\hat{\boldsymbol{g}}(\beta^{\mathrm{T}}X;\beta) - \hat{\boldsymbol{g}}(\beta_0^{\mathrm{T}}X;\beta_0)] - [\boldsymbol{g}_0(\beta^{\mathrm{T}}X;\beta) - \boldsymbol{g}_0(\beta_0^{\mathrm{T}}X)]\right\}^{\mathrm{T}}Z||\\ \leqslant& c||\hat{\dot{\boldsymbol{g}}} - \dot{\boldsymbol{g}}_0||_{\mathcal{G}}||\beta^{(r)} - \beta_0^{(r)}||[||X - E(X|\beta_0^{\mathrm{T}}X)||](||Z||),\end{aligned}$$

$$||\psi(\beta^{(r)})|| \leqslant c||\dot{\boldsymbol{g}}_0(\beta^{\mathrm{T}}X)||(||Z||)(||X||).$$

因此

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^{(r)}} ||M_{31}(\beta^{(r)})|| = o(n^{-1/2}).$$

类似地, 可以得到

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^{(r)}} ||M_{32}(\beta^{(r)})|| = o(n^{-1/2}).$$

因此 (7.5.13) 式得证.

最后证明 (7.5.14) 式. 用 $f_0(u)$ 表示 $\beta_0^{\mathrm{T}}X$ 的密度函数. 由 (7.5.6) 式可得: 对 $u \in \mathcal{U}_w$ 和 $\beta \in \mathcal{B}_n^{(r)}$ 一致成立

$$\hat{\boldsymbol{g}}(u;\beta) - \boldsymbol{g}_0(u) = [f(u)]^{-1}D^{-1}(u)\xi_{n,0}^*(u;\beta) + \frac{1}{2}h^2\mu_2\ddot{\boldsymbol{g}}_0(u) + o_P(c_n), \qquad (7.5.16)$$

其中 $c_n = n^{-1/2} + h^2$, $D(u)$ 在条件 (C7.2.6) 中定义. 由 (7.5.10) 和 (7.5.16) 式, 并使用控制收敛定理 (Loève, 2000), 可得

$$
\begin{aligned}
&\varpi(\boldsymbol{g}_0(\beta_0^{\mathrm{T}}X), \beta_0^{(r)})[\hat{\boldsymbol{g}}(\beta_0^{\mathrm{T}}X;\beta_0) - g_0(\beta_0^{\mathrm{T}}X)]\\
=& -\int C(u)[\hat{\boldsymbol{g}}(u;\beta_0) - \boldsymbol{g}_0(u)]f_0(u)\mathrm{d}u\\
=& -\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i C(\beta_0^{\mathrm{T}}X_i)D^{-1}(\beta_0^{\mathrm{T}}X_i)Z_i - \frac{1}{2}h^2\mu_2 E[C(\beta_0^{\mathrm{T}}X)\ddot{\boldsymbol{g}}_0(\beta_0^{\mathrm{T}}X)] + o_P(c_n),
\end{aligned}
$$

其中 $C(\cdot)$ 和 μ_2 分别在条件 (C7.2.6) 和定理 7.2.3 中定义. 结合 (7.5.8) 式和 $M_4(\beta_0^{(r)})$ 的定义可得

$$
M_4(\beta_0^{(r)}) = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\zeta_i - \frac{1}{2}h^2\mu_2 E[C(\beta_0^{\mathrm{T}}X)\ddot{\boldsymbol{g}}_0(\beta_0^{\mathrm{T}}X)] + o_P(c_n), \tag{7.5.17}
$$

其中 $\zeta_i = V_i - C(\beta_0^{\mathrm{T}}X_i)D^{-1}(\beta_0^{\mathrm{T}}X_i)Z_i$, $V_i = J_{\beta_0^{(r)}}^{\mathrm{T}} X_i \dot{\boldsymbol{g}}_0^{\mathrm{T}}(\beta_0^{\mathrm{T}}X_i)Z_i w(\beta_0^{\mathrm{T}}X_i)$. 因此, 对 (7.5.17) 式利用独立变量之和的中心极限定理 (Rio, 1995) 和 Slutsky 定理可得

$$
\sqrt{n}M_4(\beta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\zeta_i + o_P(1) \xrightarrow{D} N(0, \sigma^2 A).
$$

这就证明了 (7.5.14) 式. 至此我们完成了引理 7.5.2 的证明.

定理 7.2.2 的证 使用引理 7.5.2 的记号, 可将 $Q_n(\beta^{(r)})$ 分解为

$$
Q_n(\beta^{(r)}) = M_1(\beta^{(r)}) + M_2(\beta^{(r)}) + M_3(\beta^{(r)}) + M_4(\beta_0^{(r)}) + Q(\boldsymbol{g}_0, \beta^{(r)}). \tag{7.5.18}
$$

根据 $Q(\boldsymbol{g}_0, \beta_0^{(r)}) = 0$, 使用 Taylor 公式可得: 对 $\beta^{(r)} \in \mathcal{B}_n^{(r)}$ 一致成立

$$
Q(\boldsymbol{g}_0, \beta^{(r)}) = -B(\beta^{(r)} - \beta_0^{(r)}) + o_P(n^{-1/2}). \tag{7.5.19}
$$

由 (7.5.18) 和 (7.5.19) 式以及引理 7.5.2 的 (7.5.11)~(7.5.13) 式可得: 对 $\beta^{(r)} \in \mathcal{B}_n^{(r)}$ 一致成立

$$
Q_n(\hat{\boldsymbol{g}}, \beta^{(r)}) = M_4(\beta_0^{(r)}) - B(\beta^{(r)} - \beta_0^{(r)}) + o_P(n^{-1/2}). \tag{7.5.20}
$$

注意到 $Q_n(\hat{\boldsymbol{g}}, \hat{\beta}^{(r)}) = 0$, 由上式可得

$$
\hat{\beta}^{(r)} - \beta_0^{(r)} = B^{-1}M_4(\beta_0^{(r)}) + o_P(n^{-1/2}). \tag{7.5.21}
$$

由 (7.5.21) 和 (3.10.41) 式可得

$$
\sqrt{n}(\hat{\beta} - \beta_0) = \sqrt{n}\, J_{\beta_0^{(r)}} B^{-1} M_4(\beta_0^{(r)}) + o_P(1).
$$

结合上式和引理 7.5.2 即完成了定理 7.2.2 的证明.

定理 7.2.3 的证　注意到 $\hat{\beta}$ 是 $\sqrt{n}$ 相合的, 可以证明

$$\xi_{n,0}^{*}(u;\hat{\beta})=\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i Z_i K_h(\beta_0^{\mathrm{T}}X_i-u)+O_P(n^{-1/2}),$$

其中 $\xi_{n,0}^{*}(u;\hat{\beta})$ 在 (7.5.16) 式中定义. 利用 (Masry and Tjøstheim, 1995) 的定理 4.4 可得

$$\sqrt{nh}\xi_{n,0}^{*}(u;\hat{\beta})\xrightarrow{D}N\left(0,f(u)\nu_0\sigma^2D(u)\right). \tag{7.5.22}$$

因此, 由 (7.5.16) 和 (7.5.22) 式即可证得定理 7.2.3.

定理 7.2.4 的证　分解 $\hat{\sigma}^2$ 可得

$$\begin{aligned}\hat{\sigma}^2&=\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2+\frac{2}{n}\sum_{i=1}^{n}\left[\boldsymbol{g}_0(X_i^{\mathrm{T}}\beta_0)-\hat{\boldsymbol{g}}(X_i^{\mathrm{T}}\hat{\beta};\hat{\beta})\right]^{\mathrm{T}}Z_i\varepsilon_i\\&\quad+\frac{1}{n}\sum_{i=1}^{n}\left[[\boldsymbol{g}_0(X_i^{\mathrm{T}}\beta_0)-\hat{\boldsymbol{g}}(X_i^{\mathrm{T}}\hat{\beta};\hat{\beta})]^{\mathrm{T}}Z_i\right]^2\\&\equiv I_1+I_2+I_3.\end{aligned}$$

使用中心极限定理可得

$$I_1-\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i^2-\sigma^2)=O_P\left(n^{-1/2}\right).$$

由定理 7.2.1 和定理 7.2.2 可得

$$\begin{aligned}|I_2|&\leqslant\left[\sup_{(u,\beta)\in(\mathcal{U}_w,\mathcal{B}_n)}\|\hat{\boldsymbol{g}}(u;\beta)-\boldsymbol{g}_0(u)\|+O_P(n^{-1/2})\right]\cdot\frac{1}{n}\sum_{i=1}^{n}\|Z_i\|^2|\varepsilon_i|\\&=O_P\left(\left[\frac{\log n}{nh}\right]^{1/2}+h^2\right).\end{aligned}$$

同理, $I_3=O_P\left([\log n/(nh)]^{1/2}+h^2\right)$. 结合上面的结果即完成了定理 7.2.4 的证明.

定理 7.3.1 的证　利用 (7.5.20) 式和引理 7.5.2 的 (7.5.14) 式可得到

$$Q_n(\hat{\boldsymbol{g}},\beta_0^{(r)})=O_P(n^{-1/2}). \tag{7.5.23}$$

其中 $Q_n(\hat{\boldsymbol{g}},\beta_0^{(r)})$ 在 (7.5.8) 式中定义. 类似于引理 3.12.2～ 引理 3.12.4 的证明不难证得

$$\frac{1}{n}\sum_{i=1}^{n}\hat{Z}_i(\beta_0^{(r)})\hat{Z}_i^{\mathrm{T}}(\beta_0^{(r)})\xrightarrow{P}A_0, \tag{7.5.24}$$

$$\max_{1\leqslant i\leqslant n}||\hat{Z}_i(\beta_0^{(r)})|| = o_P(n^{1/2}), \tag{7.5.25}$$

$$\lambda(\beta_0^{(r)}) = o_P(n^{-1/2}), \tag{7.5.26}$$

其中 A_0 在定理 7.3.1 中定义, $\hat{Z}_i(\beta^{(r)})$ 在 (7.3.1) 中定义.

利用 (7.5.23)~(7.5.26) 式并类似于定理 3.12.1 的证明即可完成定理 7.3.1 的证明.

定理 7.3.2 的证明类似于定理 3.12.2 的证明, 这里证明省略.

参 考 文 献

Arnold S F. 1981. The Theory of Linear Models and Multivariate Analysis. New York: John Wiley & Sons

Cai Z W, Fan J, Yao Q. 2000. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association, 95: 941~956

Carroll R J, Fan J, Gijbels I, et al.. 1997. Generalized partially linear single index models. Journal of the American Statistical Association, 92: 477~489

Cui X, Härdle W K, Zhu L X. 2011. The EFM approach for single-index models. The Annals of Statistics, 39(3): 1658~1688

Doukhan P, Massart P, Rio E. 1995. Invariance principles for absolutely regular empirical processes. Annals of Institute H Poincaré Probability Statistics, 31: 393~427

Fan J, Gijbels I. 1996. Local Polynomial Modeling and Its Applications. London: Chapman and Hall

Fan J, Yao Q, Cai Z W. 2003. Adaptive varying-coefficient linear models. Journal of the Royal Statistical Society, Series B, 65: 57~80

Härdle W, Hall P, Ichimura H. 1993. Optimal smoothing in single-index models. The Annals of Statistics, 21: 157~178

Loève M. 2000. Probability Theory I, 4th Edition. New York: Springer-Verlag

Lu Z D, Tjøstheim D, Yao Q W. 2007. Adaptive varying-coefficient linear models for stochastic processes: Asymptotic theory. Statistica Sinica 17: 177~197

Masry E, Tjøstheim D. 1995. Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. Econometric Theory, 11: 258~289

Rio E. 1995. The functional law of the iterated logarithm for stationary strongly mixing sequence. The Annals of Probability, 23: 1188~1203

Ruppert D, Wand M P, Carroll R J. 2003. Semiparametric Regression. London: Cambridge University Press

Stute W. 1982. A law for the logarithm of kernel density estimators. The Annals of Probability, 10: 414～422

Wang J L, Xue L G, Zhu L X, et al.. 2010. Estimation for a partial-linear single-index model. The Annals of Statistics, 38: 246～274

Wu T Z, Lin H Q, Yu Y. 2011. Single-index coefficient models for nonlinear time series. Journal of Nonparametric Statistics, 23(1): 37～58

Xia Y C, Li W K. 1999. On single-index coefficient regression models. Journal of the American Statistical Association, 94: 1275～1285

Xue L G, Pang Z. 2012. Statistical inference for a single-index varying-coefficient model. statistics and computing, published online, DOI: 10.1007/s11222-012-9332-x

Xue L G, Wang Q H. 2012. Empirical likelihood for single-index varying-coefficient models. Bernoulli, 18(3): 836～856

Yu Y, Ruppert D. 2002. Penalized spline estimation for partially linear single-index models. Journal of the American Statistical Association, 97: 104～105

第8章　部分变系数单指标模型

8.1　引　　言

部分线性变系数模型是流行的半参数模型之一，其结构如下：

$$Y = \theta^{\mathrm{T}}(U)Z + \beta_0^{\mathrm{T}}X + \varepsilon, \tag{8.1.1}$$

其中 $(X, Z) \in \mathbf{R}^p \times \mathbf{R}^q$, $U \in \mathbf{R}^1$ 为协变量, Y 为响应变量, $\theta(\cdot) = (\theta_1(\cdot), \cdots, \theta_q(\cdot))^{\mathrm{T}}$ 为未知函数向量, $\beta_0 = (\beta_{01}, \cdots, \beta_{0p})^{\mathrm{T}}$ 为未知参数向量, ε 是期望 0 方差为 σ^2 的随机误差.

许多学者对模型 (8.1.1) 已经进行了研究. 例如, Xia 等 (2004), Ahmad 等 (2005) 以及 Fan 和 Huang(2005) 等. 本书第 6 章阐述了模型 (8.1.1) 的理论和方法的一些研究成果. Fan 和 Huang(2005) 利用模型 (8.1.1) 分析了波士顿住房数据. 该数据集由住房价格和解释住房价格变化的七个变量构成. 住房价格是 1970 年居住在波士顿市区内的 506 个地带的物主对其住房标价的中位数. 七个协变量是: 城市人均犯罪率 (CRIM), 每个住宅的房间平均数 (RM), 每 10000 美元含全部财产税 (TAX), 一氧化氮含量 (NOX), 市区学校的师生比例 (PTRATIO), 在 1940 年之前建造的住房比例 (AGE), 底层居民的比例 (LSTAT). 为符号简单起见, 将协变量 CRIM, RM, TAX, NOX, PTRATIO 和 AGE 分别记为 $Z_2, \cdots, Z_7$.

取 $Z_1 = 1$ 为截距项, $U = \sqrt{\mathrm{LSTAT}}$. Fan 和 Huang(2005) 首先利用变系数模型

$$Y = \sum_{j=1}^{7} \theta_j(U) Z_j + \varepsilon$$

拟合给定的数据. 他们利用 Fan 等 (2001) 提出的广义似然比统计量来检验每一个变系数函数是否显著变化, 结果显示 PTRATIO 和 AGE 的系数函数在 1% 的水平下无显著变化. 因此, 他们令 PTRATIO 和 AGE 的系数为常数, 并采用部分线性变系数模型

$$Y = \sum_{j=1}^{5} \theta_j(U) Z_j + \beta_1 Z_6 + \beta_2 Z_7 + \varepsilon$$

重新拟合数据, 考虑了 β_1 和 β_2 的统计显著性, 即 “无显著变化”. 为了说明 PTRATIO 和 AGE 的系数是否具有显著的统计意义, Fan 和 Huang(2005) 采用 GLR 和

Wald 两种方法来检验, 结果显示 AGE 的系数接近于 0, 这是不切实际的. 因为这意味着 AGE 和自住型住房价格的中位数 (Y) 没有显著的线性关系. 因此, 本章采用模型

$$Y=\sum_{j=1}^{5}\theta_j(U)Z_j+g(\beta_1Z_6+\beta_2Z_7)+\varepsilon$$

来拟合这个数据, 其中 $g(\cdot)$ 为非线性函数. 基于上述模型, 我们的研究表明 PTRATIO 和 AGE 都影响住房价格, 这是一个合理的结果. 受这个实际应用问题的启发, 考虑如下形式的部分变系数单指标模型:

$$Y=\theta^{\mathrm{T}}(U)Z+g(\beta_0^{\mathrm{T}}X)+\varepsilon, \tag{8.1.2}$$

其中 $g(\cdot)$ 为未知联系函数. 考虑到模型的可识别性, 假定 $||\beta_0||=1$, 其中 $||\cdot||$ 表示 Euclid 范数. 进一步, 假设 $E(\varepsilon|U,X,Z)=0$, $\mathrm{var}(\varepsilon|U,X,Z)=\sigma^2(U)$.

在模型 (8.1.2) 中, $\theta(\cdot)$ 表明 Z 和 U 之间的一种特殊的交互作用. 因此, 它可以看作部分线性单指标模型的自然推广. 相关研究可见文献 (Carroll et al., 1997; Yu and Ruppert, 2002; Zhu and Xue, 2006) 等. 该模型具有一般性并且可以避免"维数灾祸"(curse of dimensionality). 它除包含部分线性变系数模型和部分线性单指标模型外, 还以许多其他重要模型作为特例. 例如, 若 $\beta_0=1$ 且 $\theta(\cdot)$ 为常数参数向量, 则模型 (8.1.2) 成为部分线性模型 (见第 2 章). 若模型 (8.1.2) 中 $g(\cdot)=0$, 则它变为变系数模型 (见第 5 章). 有关研究包括 (Hastie and Tibshirani, 1993; Hoover et al., 1998; Wu et al., 1998; Cai, Fan and Li, 2000; Cai, Fan and Yao, 2000; Fan and Zhang, 2000; Chiang et al., 2001; Huang et al., 2002; Xue and Zhu, 2007) 等. 若 $p=1$ 和 $\beta_0=1$, 则模型 (8.1.2) 变成一类变系数非参数模型. 模型 (8.1.2) 也包含缺少变系数分量的单指标模型 (见第 3 章). 主要研究包括 (Wahba, 1984; Härdle and Stoker, 1989;, Li, 1991; Härdle et al., 1993; Green and Silverman, 1994; Zhu and Fangm 1996; Chiou and Müller, 1998; Xia and Li, 1999; Xue and Zhu, 2006) 等.

模型 (8.1.2) 兼有单指标模型和变系数模型的特点. 已有许多估计方法用于研究单指标模型, 如切片逆回归、核光滑、局部线性方法、惩罚样条方法以及平均导数法等 (见第 3 章). 另外, 也有多种估计方法对变系数模型进行统计推断, 主要包括局部最小二乘方法、核方法、光滑样条方法、局部多项式回归、局部多项式核广义估计方程方法以及基于基函数的近似方法等 (见第 5 章). 模型 (8.1.2) 也是部分线性单指标模型的合理推广. Yu 和 Ruppert(2002) 提出 P 样条方法估计部分线性单指标模型. P 样条单指标模型中的所有参数都可由非线性惩罚最小二乘方法同时估计. 最近, Wong 等 (2008) 研究了模型 (8.1.2), 他们结合二元局部线性光滑和平均方法同时得到了模型中未知函数的拟初始估计. 然而这种方法可能会导致估计是不相合的, 因为对一元系数函数和一元非参数函数二元局部光滑方法得到两个二元

估计函数, 这显然是渐近有偏的, 因此二元估计函数的算术平均一般得不到相合估计. 当然也可使用其他现有技术和估计方法同时获得未知量的估计, 如光滑样条或多项式样条等方法. 但是这样的方法将会使研究估计量的渐近性质变得非常困难, 因为这些估计量彼此之间存在非常复杂的依赖关系, 而且它的计算量大, 尤其当 Z 的维数非常大时更是如此. 其他研究工作见 (Huang, 2011).

由于模型 (8.1.2) 关于变系数向量 $\theta(\cdot)$、非参数联系函数 $g(\cdot)$ 和未知参数向量 β_0 的结构非常复杂, 因此需要寻找一种新的途径或技术并结合现有方法来获得有关估计量. 本章运用逐步估计方法来估计模型 (8.1.2) 中的参数和非参数分量 β_0, $\theta(\cdot)$ 和 $g(\cdot)$, 并建立所提出的估计量的渐近理论. 本章的方法不需要迭代且在较弱的条件下便于实施. 其基本思想是: 首先假定单指标参数向量已知, 将模型 (8.1.2) 写作一个变系数模型; 然后运用局部线性回归方法和 Nadaraya-Watson 核估计方法获得系数函数向量的初始估计 $\check{\theta}(U;\beta_0)$; 最后, 假定 β_0 已知, 基于 $Y-Z^{\mathrm{T}}\check{\theta}(U;\beta_0)$ 关于 $\beta_0^{\mathrm{T}}X$ 的回归运用局部线性方法估计未知联系函数 $g(\cdot)$. 这个估计与未知参数 β_0 有关. 因此, 可用最小二乘方法来估计未知参数向量 β_0, 并得到 $\theta(\cdot)$ 和 $g(\cdot)$ 的最终估计. 需要指出的是, 由于模型的复杂性使得研究所提出估计量的渐近性质时具有挑战性, 需要一些特别的技巧来研究这些估计量的渐近性质.

本章结构如下：8.2 节介绍估计方法, 8.3 节给出所提出的估计量的渐近性质, 8.4 节应用主要结果构造兴趣参数的置信区间或置信域, 8.5 节考虑带宽选择问题, 8.6 节通过模拟仿真研究和实际数据分析来说明所提出的估计方法的有效性. 8.7 节给出主要结果的证明. 本章的主要内容取材于文献 (Wang and Xue, 2011).

8.2 估计方法

假设 $\{(U_i,X_i,Z_i,Y_i),1\leqslant i\leqslant n\}$ 是来自模型 (8.1.2) 的独立同分布 (iid) 观测样本, 即有

$$Y_i=\theta^{\mathrm{T}}(U_i)Z_i+g(\beta_0^{\mathrm{T}}X_i)+\varepsilon_i,\quad i=1,\cdots,n, \tag{8.2.1}$$

其中 $X_i=(X_{i1},\cdots,X_{ip})^{\mathrm{T}}$, $Z_i=(Z_{i1},\cdots,Z_{iq})^{\mathrm{T}}$, $\varepsilon_1,\cdots,\varepsilon_n$ 为 iid 误差且 $E(\varepsilon_i|U_i,X_i,Z_i)=0$, $i=1,\cdots,n$.

首先考虑 $\theta(\cdot)$ 的估计. 由 (8.2.1) 式可知

$$E(Y_i|\beta_0^{\mathrm{T}}X_i,U_i)=\theta^{\mathrm{T}}(U_i)E(Z_i|\beta_0^{\mathrm{T}}X_i,U_i)+g(\beta_0^{\mathrm{T}}X_i).$$

再结合 (8.2.1) 式, 有

$$Y_i-m_1(\beta_0^{\mathrm{T}}X_i,U_i)=\theta^{\mathrm{T}}(U_i)[Z_i-m_2(\beta_0^{\mathrm{T}}X_i,U_i)]+\varepsilon_i, \tag{8.2.2}$$

其中 $m_1(t,u)=E(Y|\beta_0^{\mathrm{T}}X=t,U=u)$, $m_2(t,u)=E(Z|\beta_0^{\mathrm{T}}X=t,U=u)$. 若 $m_1(t,u)$, $m_2(t,u)$ 和 β_0 已知, 则 (8.2.2) 式变为变系数模型. 因此, 变系数模型的估计方法可以用来估计 $\theta(\cdot)$. 然而, 这些量是未知的, 需要用它们的估计量来替代. 我们用核回归方法估计 $m_1(t,u)$ 和 $m_2(t,u)$, 因为它们是二元函数. 令集合 $\mathcal{X}$ 为使得 $\beta_0^{\mathrm{T}}X$ 的密度函数 $r(t)$ 在 $\{\beta_0^{\mathrm{T}}X:x\in\mathcal{X}\}$ 上有界且远离 0 的开凸集. $I_{\mathcal{X}}(x)$ 为 $\mathcal{X}$ 上的示性函数, 即若 $x\in\mathcal{X}$, $I_{\mathcal{X}}(x)$ 为 1, 否则为 0. 于是, 对于固定的 β_0, $m_1(t,u)$ 和 $m_2(t,u)$ 的 Nadaraya-Watson 核估计分别定义为

$$\hat{m}_1(t,u;\beta_0)=\frac{\sum_{i=1}^{n}I_{\mathcal{X}}(X_i)Y_iK_1\left(\frac{\beta_0^{\mathrm{T}}X_i-t}{h_1^{1/2}},\frac{U_i-u}{h_1^{1/2}}\right)}{\sum_{i=1}^{n}I_{\mathcal{X}}(X_i)K_1\left(\frac{\beta_0^{\mathrm{T}}X_i-t}{h_1^{1/2}},\frac{U_i-u}{h_1^{1/2}}\right)},$$

$$\hat{m}_2(t,u;\beta_0)=\frac{\sum_{i=1}^{n}I_{\mathcal{X}}(X_i)Z_iK_1\left(\frac{\beta_0^{\mathrm{T}}X_i-t}{h_1^{1/2}},\frac{U_i-u}{h_1^{1/2}}\right)}{\sum_{i=1}^{n}I_{\mathcal{X}}(X_i)K_1\left(\frac{\beta_0^{\mathrm{T}}X_i-t}{h_1^{1/2}},\frac{U_i-u}{h_1^{1/2}}\right)},$$

其中 $K_1(\cdot,\cdot)$ 为 $\mathbf{R}^2$ 上的核函数, $h_1=h_1(n)$ 为带宽, 且 $0<h_1<1$, $h_1\to 0$.

下面应用局部线性回归方法来估计系数函数 $\{\theta_j(\cdot),1\leqslant j\leqslant q\}$(Fan and Gijbels, 1996). 当 U 在 u 的一个小邻域内时, 可以用一个线性函数局部地逼近 $\theta_j(U)$,

$$\theta_j(U)\approx\theta_j(u)+\theta_j'(u)(U-u)\equiv a_j+b_j(U-u),\quad j=1,\cdots,q.$$

$\theta_j(u)$ 和 $\theta_j'(u)$ 的初始估计分别定义为 $\check{\theta}_j(u;\beta_0)=\check{a}_j$ 和 $\check{\theta}_j'(u;\beta_0)=\check{b}_j$, 其中 $\{(\check{a}_j,\check{b}_j);j=1,\cdots,q\}$ 为下面加权平方和的极小值点:

$$\sum_{i=1}^{n}\left\{\widetilde{Y}_i-\sum_{j=1}^{q}[a_j+b_j(U_i-u)]\,\widetilde{Z}_{ij}\right\}^2K_{h_2}(U_i-u),$$

其中 $\widetilde{Y}_i=Y_i-\hat{m}_1(\beta_0^{\mathrm{T}}X_i,U_i;\beta_0)$, $\widetilde{Z}_{ij}=Z_{ij}-\hat{m}_2(\beta_0^{\mathrm{T}}X_i,U_i;\beta_0)$, $h_2=h_2(n)$ 为带宽且 $0<h_2\to 0$, $K(\cdot)$ 为核函数且 $K_{h_2}(\cdot)=h_2^{-1}K(\cdot/h_2)$. 由最小二乘理论可得

$$(\check{\theta}^{\mathrm{T}}(u;\beta_0),h_2\check{\theta}'^{\mathrm{T}}(u;\beta_0))^{\mathrm{T}}=R_n^{-1}(u;\beta_0)\eta_n(u;\beta_0),\tag{8.2.3}$$

其中 $\check{\theta}(u;\beta_0)=(\check{\theta}_1(u;\beta_0),\cdots,\check{\theta}_q(u;\beta_0))^{\mathrm{T}}$, $\check{\theta}'(u;\beta_0)=(\check{\theta}_1'(u;\beta_0),\cdots,\check{\theta}_q'(u;\beta_0))^{\mathrm{T}}$, $\eta_n(u;\beta_0)=(\eta_{n,0}^{\mathrm{T}}(u;\beta_0),\eta_{n,1}^{\mathrm{T}}(u;\beta_0))^{\mathrm{T}}$,

$$\eta_{n,j}(u;\beta_0)=\frac{1}{n}\sum_{i=1}^{n}\widetilde{Z}_i\left(\frac{U_i-u}{h_2}\right)^jK_{h_2}(U_i-u)\widetilde{Y}_i,\quad j=0,1,\tag{8.2.4}$$

$$R_n(u;\beta_0) = \begin{pmatrix} R_{n,0}(u;\beta_0) & R_{n,1}(u;\beta_0) \\ R_{n,1}(u;\beta_0) & R_{n,2}(u;\beta_0) \end{pmatrix},$$

$$R_{n,j}(u;\beta_0) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{Z}_i \widetilde{Z}_i^{\mathrm{T}} \left(\frac{U_i - u}{h_2}\right)^j K_{h_2}(U_i - u), \quad j = 0,1,2,3. \tag{8.2.5}$$

基于 $\check{\theta}(u;\beta_0)$, 可以通过 $Y - Z^{\mathrm{T}}\check{\theta}(U;\beta_0)$ 对于 $\beta_0^{\mathrm{T}}X$ 的局部线性光滑来定义 $g(\cdot)$ 和 $g'(\cdot)$ 的初始估计. 给定 $h_3 = h_3(n)$, $0 < h_3 \to 0$ 和核函数 $K(\cdot)$, $K_{h_3}(\cdot) = h_3^{-1}K(\cdot/h_3)$, 关于 d_ν 极小化加权平方和

$$\sum_{i=1}^{n} I_{\mathcal{X}}(X_i)\left[Y_i - \check{\theta}^{\mathrm{T}}(U_i;\beta_0)Z_i - d_0 - d_1(\beta_0^{\mathrm{T}}X_i - t)\right]^2 K_{h_3}(\beta_0^{\mathrm{T}}X_i - t),$$

可以得到极小值点 $\check{d}_\nu$, $\nu = 0,1$. 对于固定的 β_0, $g(\cdot)$ 和 $g'(\cdot)$ 的估计可以分别定义为 $\check{g}(t;\beta_0) = \check{d}_0$ 和 $\check{g}'(t;\beta_0) = \check{d}_1$. 简单计算可得

$$\check{g}(t;\beta_0) = \sum_{i=1}^{n} W_{ni}(t,\beta_0)\left[Y_i - \check{\theta}^{\mathrm{T}}(U_i;\beta_0)Z_i\right], \tag{8.2.6}$$

其中

$$W_{ni}(t;\beta_0) = \frac{n^{-1}I_{\mathcal{X}}(X_i)K_{h_3}(\beta_0^{\mathrm{T}}X_i - t)[A_{n,2}(t;\beta_0) - (\beta_0^{\mathrm{T}}X_i - t)A_{n,1}(t;\beta_0)]}{A_{n,0}(t;\beta_0)A_{n,2}(t;\beta_0) - A_{n,1}^2(t;\beta_0)}, \tag{8.2.7}$$

$$A_{n,l}(t;\beta_0) = \frac{1}{n}\sum_{i=1}^{n} I_{\mathcal{X}}(X_i)(\beta_0^{\mathrm{T}}X_i - t)^l K_{h_3}(\beta_0^{\mathrm{T}}X_i - t), \quad l = 0,1,2. \tag{8.2.8}$$

上述 $\theta(\cdot)$ 和 $g(\cdot)$ 的估计都是在假定 β_0 已知时得到的. 下面来估计 β_0. 极小化误差平方和

$$Q(\beta) = \sum_{i=1}^{n} I_{\mathcal{X}}(X_i)\left[Y_i - Z_i^{\mathrm{T}}\check{\theta}(U_i;\beta) - \check{g}(\beta^{\mathrm{T}}X_i;\beta)\right]^2, \tag{8.2.9}$$

可以得到 β_0 的最终估计量 $\hat{\beta}$. 利用 $\hat{\beta}$, 可以定义 $\theta(u)$, $\theta'(u)$ 和 $g(t)$ 的最终估计量为 $\hat{\theta}(u) = \check{\theta}(u;\hat{\beta})$, $\hat{\theta}'(u) = \check{\theta}'(u;\hat{\beta})$ 和

$$\hat{g}(t) \equiv \check{g}(t;\hat{\beta}) = \sum_{i=1}^{n} W_{ni}(t;\hat{\beta})\left[Y_i - \hat{\theta}^{\mathrm{T}}(U_i)Z_i\right]. \tag{8.2.10}$$

于是, $\sigma^2(u)$ 的估计量 $\hat{\sigma}^2(u)$ 可以定义为

$$\hat{\sigma}^2(u) = \frac{1}{n\hat{f}(u)}\sum_{i=1}^{n}[Y_i - \hat{\theta}^{\mathrm{T}}(U_i)Z_i - \hat{g}(\hat{\beta}^{\mathrm{T}}X_i)]^2 K_{h_2}(U_i - u), \tag{8.2.11}$$

其中

$$\hat{f}(u) = \frac{1}{n}\sum_{i=1}^{n} K_{h_2}(U_i - u). \tag{8.2.12}$$

注意到, 假定 $||\beta_0|| = 1$, 这个约束条件可以用来提高 β_0 的估计的有效性. 为此, 我们采用 Yu 与 Ruppert(2002) 和 Zhu 与 Xue(2006) 提出的 “去一分量” 方法. 其具体思路表述如下. 假定真实参数 β_0 包含一个正的分量 (否则, 考虑 $-\beta_0$). 不失一般性, 假定 $\beta_{r0} > 0$, 其中 β_{r0} 是 $\beta_0 = (\beta_{10}, \cdots, \beta_{p0})^{\mathrm{T}}$ 的第 r 个分量, $1 \leqslant r \leqslant p$. 对 $\beta = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$, 令 $\beta^{(r)} = (\beta_1, \cdots, \beta_{r-1}, \beta_{r+1}, \cdots, \beta_p)^{\mathrm{T}}$ 是去掉 β 的第 r 个分量 β_r 后得到的 $p-1$ 维参数向量, 则有

$$\beta = \beta(\beta^{(r)}) = \left(\beta_1, \cdots, \beta_{r-1}, (1 - ||\beta^{(r)}||^2)^{1/2}, \beta_{r+1}, \cdots, \beta_p\right)^{\mathrm{T}}. \tag{8.2.13}$$

真实参数 $\beta_0^{(r)}$ 满足 $||\beta_0^{(r)}|| < 1$. 因此, β 在 $\beta_0^{(r)}$ 的邻域内无限可微, 其Jacobian 矩阵为

$$J_{\beta^{(r)}} = \frac{\partial \beta}{\partial \beta^{(r)}} = (\gamma_1, \cdots, \gamma_p)^{\mathrm{T}}, \tag{8.2.14}$$

其中 $\gamma_s(1 \leqslant s \leqslant p, s \neq r)$ 是 $p-1$ 维单位向量, 其第 $s - I(s > \gamma)$ 个分量为 1, 且 $\gamma_r = -(1 - ||\beta^{(r)}||^2)^{-1/2}\beta^{(r)}$.

注意到, $Q(\beta) = Q(\beta(\beta^{(r)})) \equiv \widetilde{Q}(\beta^{(r)})$, 我们可以通过极小化 $\widetilde{Q}(\beta^{(r)})$ 得到 $\beta_0^{(r)}$ 的估计量 $\hat{\beta}^{(r)}$, 从而通过简单的变换得到 β_0 的估计量. 求解方程

$$\sum_{i=1}^{n} I_{\mathcal{X}}(X_i)\left[Y_i - \check{\theta}^{\mathrm{T}}(U_i;\beta)Z_i - \check{g}(\beta^{\mathrm{T}}X_i;\beta)\right]\check{g}'(\beta^{\mathrm{T}}X_i;\beta)J_{\beta^{(r)}}^{\mathrm{T}}X_i = 0, \tag{8.2.15}$$

可以得到估计量 $\hat{\beta}^{(r)}$, 其中 $\check{\theta}(U_i;\beta)$ 和 $\check{g}(\beta^{\mathrm{T}}X_i;\beta)$ 分别由 (8.2.3) 和 (8.2.6) 式定义,

$$\check{g}'(t;\beta) = \sum_{i=1}^{n} \widetilde{W}_{ni}(t,\beta)\left[Y_i - \check{\theta}^{\mathrm{T}}(U_i;\beta)Z_i\right],$$

$$\widetilde{W}_{ni}(t;\beta) = \frac{n^{-1}I_{\mathcal{X}}(X_i)K_{h_3}(\beta^{\mathrm{T}}X_i - t)[(\beta^{\mathrm{T}}X_i - t)A_{n,0}(t;\beta) - A_{n,1}(t;\beta)]}{A_{n,0}(t;\beta)A_{n,2}(t;\beta) - A_{n,1}^2(t;\beta)},$$

$\{A_{n,l}(t;\beta), l = 0, 1, 2\}$ 由 (8.2.8) 式定义.

我们可以用 Newton-Raphson 算法的 Fisher 得分法来求解估计方程 (8.2.15). 对于模型 (8.1.2), 该算法收敛速度非常快, 即使初始值较差也是如此.

8.3　渐近性质

令 $\mathcal{B} = \{\beta \in \mathbf{R}^p : ||\beta|| = 1, \beta_r > 0\}$, 则 β_0 是紧集 $\mathcal{B}$ 的内点. 对任一正常数 C, 记 $\mathcal{B}_n = \{\beta \in \mathcal{B} : ||\beta - \beta_0|| \leqslant Cn^{-1/2}\}$. 此定义基于如下事实, 我们期望 $\hat{\beta}$ 是 $\sqrt{n}$ 相

合的, 因此应该在 β 距 β_0 的阶为 $n^{-1/2}$ 范围内寻找 $Q(\beta)$ 的极小值点, 详见定理 8.3.3 的证明. Härdle 等 (1993) 在研究单指标模型时也作过类似的限制. 另外, 我们也可以沿用 Xia 等 (2007) 的思想, 使用高阶局部光滑来获得 $\sqrt{n}$ 相合的初始估计.

为得到本章的渐近性质, 我们首先给出下列正则条件.

(C8.3.1) 集合 $\mathcal{X}$ 是开凸集, 且使得对任何接近 β_0 的点 β, 变量 $\beta^{\mathrm{T}}X$ 的密度函数 $r(t)$ 在 $\mathcal{T}_\beta=\{x^{\mathrm{T}}\beta: x\in\mathcal{X}\}$ 上是 Lipschitz 连续的且远离 0.

(C8.3.2) $g(t)$, $\mu_{1j}(t)$ 和 $\mu_{2j}(t)$ 在 $\mathcal{T}_{\beta_0}$ 上有有界的二阶连续导数, 其中 $\mu_{1j}(t)$ 和 $\mu_{2j}(t)$ 分别为 $\mu_1(t)=E(X|\beta_0^{\mathrm{T}}X=t)$ 和 $\mu_2(t)=E(Z|\beta_0^{\mathrm{T}}X=t)$ 的第 j 个分量.

(C8.3.3) U 的密度函数 $f(u)$ 在其支撑的内点 u_0 的某个邻域 $\mathcal{N}(u_0)$ 内连续且 $f(u_0)>0$. $\theta_j(\cdot)$ 在 $\mathcal{N}(u_0)$ 上有二阶连续导数, $j=1,\cdots,q$.

(C8.3.4) $(\beta_0^{\mathrm{T}}X,U)$ 的联合密度函数 $f(t,u)$ 在 $\mathcal{T}_{\beta_0}\times\mathcal{N}(u_0)$ 上有界且远离 0, 其中 $\mathcal{N}(u_0)$ 在 (C8.3.3) 中定义. $f(t,u)$, $m_1(t,u)$ 和 $m_{2j}(t,u)$ 有有界的 4 阶偏导数, 其中 $m_{2j}(t,u)$ 是 $m_2(t,u)$ 的第 j 个分量, $j=1,\cdots,q$.

(C8.3.5) 对常数 $s>2$ 和 $j=1,\cdots,q$, $E(|Z_{1j}|^{2s}|U=u)$, $E(|\varepsilon|^{2s}|U=u)$, $E\big(|Z_{1j}|^s|\beta_0^{\mathrm{T}}X=t,U=u\big)$ 和 $E(|\varepsilon|^s|X=x,U=u,Z=z)$ 是有界的.

(C8.3.6) 对某个 $\epsilon<2-s^{-1}$, $\nu=2,3$, 有 $n^{2\epsilon-1}h_\nu\to\infty$, 其中 s 如同 (C8.3.5).

(C8.3.7) 核函数 $K(\cdot)$ 为满足 Lipschitz 条件的有界对称概率密度核函数且具有有界支撑 $[-1,1]$.

(C8.3.8) $K_1(\cdot,\cdot)$ 为有界变差的右连续的 4 阶核函数且具有支撑 $[-1,1]^2$.

(C8.3.9) 矩阵 $\Omega(u_0)$ 是正定的, $\sigma^2(u)$ 和 $\Omega(u)$ 的元素在 u_0 点是连续的且 $\sigma^2(u_0)\neq 0$, 其中 $\Omega(u_0)=E\big\{[Z-E(Z|\beta_0^{\mathrm{T}}X,U)]^{\otimes 2}|U=u_0\big\}$, $A^{\otimes 2}=AA^{\mathrm{T}}$.

(C8.3.10) 矩阵 $V=E\big\{[I_{\mathcal{X}}(X)g'(\beta_0^{\mathrm{T}}X)J^{\mathrm{T}}_{\beta_0^{(r)}}(X-E(X|\beta_0^{\mathrm{T}}X))]^{\otimes 2}\big\}$ 是正定的, 其中 $J_{\beta_0^{(r)}}$ 由 (8.2.14) 式定义.

下面的定理给出了 8.2 节所得估计量的渐近性质. 首先研究 $\check{\theta}(u_0;\beta)$ 的渐近正态性.

定理 8.3.1 假设条件(C8.3.1)~(C8.3.9)成立. 如果 $nh_1^2/(h_2\log^2 n)\to\infty$, $nh_1h_2\to\infty$, $h_1^4/h_2\to 0$ 和 $nh_2^5=O(1)$, 则对任意的 $\beta\in\mathcal{B}_n$, 有

$$\sqrt{nh_2}\left[\check{\theta}(u_0;\beta)-\theta(u_0)-\frac{1}{2}h_2^2\nu_2\theta''(u_0)\right]\xrightarrow{D}N(0,\Sigma(u_0)),$$

其中 $\nu_2=\int u^2K(u)\mathrm{d}u$,

$$\Sigma(u_0)=\nu_0f^{-1}(u_0)\sigma^2(u_0)\Omega^{-1}(u_0),$$

$\nu_0 = \int K^2(u)\mathrm{d}u$, $\Omega(u_0)$ 由条件 (C8.3.9) 定义.

注意到定理 8.3.1 对任意的 $\beta \in \mathcal{B}_n$ 都成立, 因此可以证明, 若 β 用其 $\sqrt{n}$ 相合估计代替, 则定理 8.3.1 仍成立. 下文定理 8.3.3 表明 $\hat{\beta}$ 就是这样一个估计量.

注 8.3.1 定理 8.3.1 表明: 初始估计量 $\check{\theta}_j(u_0;\beta)$ 的渐近偏差是 $\frac{1}{2}h_2^2\nu_2\theta_j''(u_0)$, 渐近方差是 $(nh_2)^{-1}e_{j,q}^{\mathrm{T}}\Sigma(u_0)e_{j,q}$, 其中 $e_{j,q}$ 是第 j 个元素为 1 的 $q\times 1$ 单位向量. 因此, 可以通过极小化偏差平方与方差之和来得到估计 $\theta_j(\cdot)$ 的最优带宽 $h_{2,\mathrm{opt}}$. 从而

$$h_{2,\mathrm{opt}} = \left\{[\nu_2\theta_j''(u_0)]^{-2}e_{j,q}^{\mathrm{T}}\Sigma(u_0)e_{j,q}\right\}^{1/5} n^{-1/5}.$$

注 8.3.2 若 $g(\cdot)=0$, 则 $\Sigma(u_0)$ 成为变系数模型估计的渐近方差 (Cai, Fan and Li, 2000; Cai, Fan and Yao, 2000). 这说明模型 (8.1.2) 可适用于变系数模型.

下列定理 8.3.2 给出了 $\check{\theta}(u;\beta)$ 的一致收敛速度.

定理 8.3.2 假设定理 8.3.1 的条件成立, 又设 $\Omega(u)$ 的每个元素都满足 1 阶 Lipschitz 条件, 则

$$\sup_{u\in\mathcal{N}(u_0),\beta\in\mathcal{B}_n} ||\check{\theta}(u;\beta)-\theta(u)|| = O_P\left(\left[\frac{\log n}{nh_1}\right]^{1/2} + \left[\frac{\log n}{nh_2}\right]^{1/2} + h_1^2 + h_2^2\right).$$

其中 $\mathcal{B}_n$ 由本节第一段定义, $\mathcal{N}(u_0)$ 由条件 (C8.3.3) 定义.

定理 8.3.3 陈述了 $\hat{\beta}$ 的渐近正态性.

定理 8.3.3 假设条件 (C8.3.10) 和定理 8.3.2 的条件都成立. 进一步设 $nh_3^3/\log n\to\infty$, $nh_3^4\to 0$. 如果 β_0 的第 r 个分量为正, 则

$$\sqrt{n}(\hat{\beta}-\beta_0) \xrightarrow{D} N(0, J_{\beta_0^{(r)}}V^{-1}SV^{-1}J_{\beta_0^{(r)}}^{\mathrm{T}}),$$

其中 $J_{\beta_0^{(r)}}$ 由 (8.2.14) 式定义, V 由条件(C8.3.10)定义, 且

$$S = E\{I_{\mathcal{X}}(X)\varepsilon g'(\beta_0^{\mathrm{T}}X)J_{\beta_0^{(r)}}^{\mathrm{T}}[X-E(X|\beta_0^{\mathrm{T}}X)]\}^{\otimes 2}.$$

注 8.3.3 由 Härdle 等 (1993) 和 Carroll 等 (1997) 可知, 部分线性单指标模型中参数 β_0 的估计 $\hat{\beta}$ 的渐近方差为 S_1 的广义逆 S_1^-, 其中

$$S_1 = \sigma^{-2}E\{I_{\mathcal{X}}(X)g'(\beta_0^{\mathrm{T}}X)[X-E(X|\beta_0^{\mathrm{T}}X)]\}^{\otimes 2}.$$

注意到 S_1 有无穷多个逆矩阵, 但只有唯一的一个广义逆与 Jacobi 矩阵 $J_{\beta_0^{(r)}}$ 有关. 可以证明 $S_1^- - J_{\beta_0^{(r)}}V^{-1}SV^{-1}J_{\beta_0^{(r)}}^{\mathrm{T}}$ 是非负定矩阵, 因此当模型 (8.1.2) 变为部分线性单指标模型时定理 8.3.3 的渐近协方差矩阵在非负定意义下小于 S_1^-, 例如 (Wang et al., 2010) 中的定理 3. 因此, 本章所定义的估计量比 (Härdle et al., 1993; Carroll

et al., 1997) 中的估计量都渐近有效. 从而, 部分线性单指标模型的现有估计是可以改进的. 这也蕴涵模型 (8.1.2) 可以作为单指标模型或部分线性单指标模型来应用.

下列推论 8.3.1 是定理 8.3.1 和定理 8.3.3 的直接结果, 推论 8.3.2 是定理 8.3.2 和定理 8.3.3 的直接结果.

推论 8.3.1 在定理 8.3.3 的条件下, 有

$$\sqrt{nh_2}\left[\hat{\theta}(u_0)-\theta(u_0)-\frac{1}{2}h_2^2\nu_2\theta''(u_0)\right]\xrightarrow{D}N(0,\Sigma(u_0)),$$

其中 ν_2 和 $\Sigma(u_0)$ 在定理 8.3.1 中定义.

推论 8.3.2 在定理 8.3.3 的假设条件下, 如果对某个常数 $c>0$ 有 $h_1=h_2=cn^{-1/5}$, 则

$$\sup_{u\in\mathcal{N}(u_0)}||\hat{\theta}(u)-\theta(u)||=O_P\left(n^{-2/5}(\log n)^{1/2}\right).$$

由定理 8.3.3, 可以得到 $\hat{\beta}$ 和 β_0 之间的夹角的渐近结果.

推论 8.3.3 在定理 8.3.3 的假设条件下, 有

$$\hat{\beta}^{\mathrm{T}}\beta_0-1=O_P(n^{-1/2}),$$

其中 $\hat{\beta}^{\mathrm{T}}\beta_0$ 表示向量 $\hat{\beta}$ 和 β_0 的内积.

下列定理和推论给出估计量 $\hat{g}(x^{\mathrm{T}}\beta)$ 在 $\mathcal{X}\times\mathcal{B}_n$ 上的一致收敛速度, 其中 $\mathcal{X}$ 在条件(C8.3.1)中定义, $\mathcal{B}_n$ 在本节第一段中定义.

定理 8.3.4 在定理 8.3.3 的条件下, 如果对某些常数 $c>0$ 和 $\nu=1,2$, 有 $h_\nu=cn^{-1/5}$, 则

$$\sup_{(x,\beta)\in\mathcal{X}\times\mathcal{B}_n}\left|\hat{g}(\beta^{\mathrm{T}}x)-g(\beta^{\mathrm{T}}x)\right|=O_P\left(\left[\frac{\log n}{nh_3}\right]^{1/2}+h_3^2+n^{-2/5}(\log n)^{1/2}\right).$$

对误差方差的估计量 $\hat{\sigma}^2(u)$, 有下列结果.

定理 8.3.5 在定理 8.3.3 的条件下, 有 $\hat{\sigma}^2(u_0)$ 是 $\sigma^2(u_0)$ 的相合估计. 进一步, 如果 $f(u)$ 和 $\sigma^2(u)$ 具有有界的二阶导数, 且对某个常数 $c>0$ 和 $\nu=1,2$, 有 $h_\nu=cn^{-1/5}$, 则

$$\sup_{u\in\mathcal{N}(\beta_0)}|\hat{\sigma}^2(u)-\sigma^2(u)|=O_P\left(\left[\frac{\log n}{nh_3}\right]^{1/2}+h_3^2+n^{-2/5}(\log n)^{1/2}\right).$$

本章所提出的估计量涉及两个函数 $m_1(\cdot,\cdot)$ 和 $m_2(\cdot,\cdot)$ 的核估计. 在运用核光滑技巧来估计这两个函数时通常会使用两个不同的带宽 h_1 和 h_1^*. 为了得到所提出的估计量的渐近结果, h_1 和 h_1^* 必须满足同样的条件. 这说明如果我们使用相同的带宽来估计这两个函数, 其渐近结果不会发生改变. 为研究方便, 我们将使用一个带宽来估计 $m_1(\cdot,\cdot)$ 和 $m_2(\cdot,\cdot)$. 另外, 对多元回归核估计, 一般也是取相同的带宽, 这样做是为了避免选带宽时所带来的困难. 这种处理方法在文献中经常使用.

8.4 近似置信区间/置信域

首先考虑 $\theta(u)$ 的每一个分量的逐点置信区间的构造. 为使用推论 8.3.1 来构造 $\theta_j(u_0)$ 的置信区间, 我们首先需要估计 $\theta''(u_0)$, $\sigma^2(u_0)$, $f(u_0)$ 和 $\Omega(u_0)$.

对 $j=1,\cdots,q$, 由 3 阶局部多项式拟合以及适当的带宽 $h_*=O(n^{-1/7})$—— 估计 $\theta_j''(u_0)$ 的最优带宽, 可以得到其相合估计量 $\hat{\theta}_j''(u_0)$. $\sigma^2(u_0)$ 和 $f(u_0)$ 的估计量由 (8.2.11) 和 (8.2.12) 式给出. $\Omega(u_0)$ 的估计量可以定义为

$$\hat{\Omega}(u_0)=\frac{1}{n\hat{f}(u_0)}\sum_{i=1}^{n}I_{\mathcal{X}}(X_i)[Z_i-\hat{m}_2(\hat{\beta}^{\mathrm{T}}X_i,U_i;\hat{\beta})]^{\otimes 2}K_{h_2}(U_i-u_0).$$

最后, 用 $\hat{\sigma}^2(u_0)$, $\hat{f}(u_0)$ 和 $\hat{\Omega}^{-1}(u_0)$ 替代 $\Sigma(u_0)$ 中的 $\sigma^2(u_0)$, $f(u_0)$ 和 $\Omega^{-1}(u_0)$ 可以得到相合估计 $\hat{\Sigma}(u_0)$.

定理 8.4.1 在推论 8.3.1 的假设下, 有

$$\hat{\Sigma}^{-1/2}(u_0)(nh_2)^{1/2}\left[\hat{\theta}(u_0)-\theta(u_0)-\frac{1}{2}h_2^2\hat{\theta}''(u_0)\right]\xrightarrow{D}N(0,I_q),$$

其中 I_q 为 $q\times q$ 的单位矩阵.

由定理 8.4.1, 可以定义 $\theta_j(u)(j=1,\cdots,q)$ 的近似 $1-\alpha$ 置信区间:

$$\hat{\theta}_j(u_0)-\frac{1}{2}h_2^2\hat{\theta}_j''(u_0)\pm z_{1-\alpha/2}(nh_2)^{-1/2}\hat{\sigma}^{jj}(u_0),$$

其中 $\hat{\theta}_j(u_0)$ 和 $\hat{\theta}_j''(u_0)$ 分别为 $\hat{\theta}(u_0)$ 和 $\hat{\theta}'(u_0)$ 的第 j 个分量, $\hat{\sigma}^{jj}(u_0)^2$ 是 $\hat{\Sigma}(u_0)$ 的第 (j,j) 个元素, $z_{1-\alpha/2}$ 是标准正态分布的 $1-\alpha/2$ 分位数.

类似于 Fan 和 Gijbels(1996) 的证明, 可以得到

$$\mathrm{bias}(\hat{g}(t))\approx(1/2)\nu_2h_3^2g''(t)\quad 和\quad \mathrm{var}(\hat{g}(t))\approx\nu_0\sigma^2[nhr(t)]^{-1},$$

其中 $r(t)$, ν_0 和 ν_2 分别在条件 (C8.3.1) 和定理 8.3.1 中定义. 由以上结果以及 $g''(t)$ 和 $r(t)$ 的估计量, 也可以构造 $g(t)$ 的逐点置信区间.

为使用定理 8.3.3 来构造 β_0 的置信域, 首先定义 V 和 S 的估计 $\hat{V}$ 和 $\hat{S}$, 即

$$\hat{V}=\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{g}'(\hat{\beta}^{\mathrm{T}}X_i)I_{\mathcal{X}}(X_i)J_{\hat{\beta}^{(r)}}^{\mathrm{T}}[X_i-\hat{\mu}_1(\hat{\beta}^{\mathrm{T}}X_i)]\right\}^{\otimes 2},$$

$$\hat{S}=\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\varepsilon}_i\hat{g}'(\hat{\beta}^{\mathrm{T}}X_i)I_{\mathcal{X}}(X_i)J_{\hat{\beta}^{(r)}}^{\mathrm{T}}[X_i-\hat{\mu}_1(\hat{\beta}^{\mathrm{T}}X_i)]\right\}^{\otimes 2},$$

其中 $\hat{g}'(\hat{\beta}^{\mathrm{T}}X_i)=\breve{g}'(\hat{\beta}^{\mathrm{T}}X_i;\hat{\beta})$, $\hat{\mu}_1(t)=\sum_{i=1}^{n}W_{ni}(t;\hat{\beta})X_i$ 是 $\mu_1(t)$ 的估计, $J_{\hat{\beta}^{(r)}}$ 是 $J_{\beta_0^{(r)}}$ 的估计, $\hat{\varepsilon}_i=Y_i-\hat{\theta}^{\mathrm{T}}(U_i)Z_i-\hat{g}(\hat{\beta}^{\mathrm{T}}X_i)$. 可以证明 $J_{\hat{\beta}^{(r)}}\xrightarrow{P}J_{\beta_0^{(r)}}$, $\hat{V}\xrightarrow{P}V$ 和 $\hat{S}\xrightarrow{P}S$. 由定理 8.3.3, 我们有

$$\left(n^{-1}J_{\hat{\beta}^{(r)}}\hat{V}^{-1}\hat{S}\hat{V}^{-1}J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\right)^{-1/2}(\hat{\beta}-\beta_0)\xrightarrow{D}N(0,I_p).$$

再由 (Arnold, 1981) 中的定理 10.2d 可得下面的极限分布.

定理 8.4.2 在定理 8.3.3 的假设条件下, 有

$$(\hat{\beta}-\beta_0)^{\mathrm{T}}\left(n^{-1}J_{\hat{\beta}^{(r)}}\hat{V}^{-1}\hat{S}\hat{V}^{-1}J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\right)^{-1}(\hat{\beta}-\beta_0)\xrightarrow{D}\chi_p^2.$$

由定理 8.4.2, 可以定义 β_0 的近似 $1-\alpha$ 置信域

$$R_\alpha=\left\{\beta:(\hat{\beta}-\beta)^{\mathrm{T}}\left(n^{-1}J_{\hat{\beta}^{(r)}}\hat{V}^{-1}\hat{S}\hat{V}^{-1}J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\right)^{-1}(\hat{\beta}-\beta)\leqslant\chi_p^2(1-\alpha)\right\}.$$

其中 $\chi_p^2(1-\alpha)$ 是自由度为 p 的 χ^2 分布的 $1-\alpha$ 分位数, $0<\alpha<1$.

8.5 带宽选择

在局部线性回归中, 带宽很大程度上影响估计精度, 而核函数则通常对估计精度影响较小 (Fan and Gijbels, 1996). 可以通过检查拟合曲线来主观选择估计量 $\hat{\theta}(\cdot)$ 和 $\hat{g}(\cdot)$ 的适当带宽. 然而, 一个自动选择带宽方法不仅具有理论和实际意义, 而且通常需要提供一个由数据决定的适当带宽范围的初步思路.

本章采用最小二乘交叉验证 (CV) 方法来选择估计量 $\hat{\theta}(\cdot)$ 和 $\hat{g}(\cdot)$ 的带宽. CV 统计量由下式给定:

$$\mathrm{CV}(h)=\frac{1}{n}\sum_{i=1}^{n}I_{\mathcal{X}}(X_i)\left[Y_i-Z_i^{\mathrm{T}}\hat{\theta}_h^{(-i)}(U_i)-\hat{g}_h^{(-i)}(X_i^{\mathrm{T}}\hat{\beta}^{(-i)})\right]^2,\tag{8.5.1}$$

其中 $\hat{\theta}_h^{(-i)}(\cdot)$, $\hat{g}_h^{(-i)}(\cdot)$ 和 $\hat{\beta}^{(-i)}$ 分别为去掉第 i 个观测值之后的估计量 $\hat{\theta}(\cdot)$, $\hat{g}(\cdot)$ 和 $\hat{\beta}$, 且 h 代替 h_1, h_2 和 h_3. 可以通过极小化 (8.5.1) 式来选择 CV 带宽 h_{CV}, 即 $h_{\mathrm{CV}}=\min\limits_{h>0}\mathrm{CV}(h)$. 我们取 $h_1=h_2=h_{\mathrm{CV}}$ 来计算估计量 $\hat{m}_1(\cdot,\cdot)$, $\hat{m}_2(\cdot,\cdot)$ 和 $\hat{\theta}(\cdot)$. 在计算 $\hat{g}(\cdot)$ 和 $\hat{\beta}$ 时, 取带宽

$$h_3=h_{\mathrm{CV}}n^{1/5}n^{-1/4}(\log n)^{-1/20}=h_{\mathrm{CV}}(n\log n)^{-1/20}.\tag{8.5.2}$$

需要注意的是, $\hat{\theta}(\cdot)$ 的 $(nh_2)^{1/2}$ 渐近正态性蕴涵定理 8.3.1 中 h_1 和 h_2 的适当选择仅依赖均方误差的二阶或高阶项. 这表明就它的均方误差而言, h_1 和 h_2 的选

择对 $\hat{\theta}(\cdot)$ 并不起决定作用. 因此, 可以通过极小化 (8.5.1) 式得到 h_1 和 h_2 的 CV 带宽. 类似地, 适当选择 h_3 会影响 $\hat{g}(\cdot)$ 的均方误差的二阶或高阶项. 因此, 可以使用由 (8.5.2) 式定义的带宽 h_3.

8.6 数值计算结果

本节通过模拟研究和实际数据分析来评估 8.2 节所提出的估计方法.

8.6.1 模拟研究

我们首先通过模拟研究评估所提出的估计方法在有限样本下的性能. 考虑模型

$$Y = \theta_1(U)Z_1 + \theta_2(U)Z_2 + g(\beta_1 X_1 + \beta_2 X_2) + \varepsilon,$$

其中 $(\beta_1, \beta_2)^{\mathrm{T}} = (1/2, \sqrt{3}/2)^{\mathrm{T}}$, $U \sim U[0,1]$, $(X_1, X_2)^{\mathrm{T}}$ 是二维随机向量, 其分量独立且服从 $U[-0.5, 0.5]$ 分布, Z_1 和 Z_2 均服从标准正态分布, $\varepsilon \sim N(0, 0.3^2)$, $\theta_1(u) = 4\cos(2\pi u)$, $\theta_2(u) = 12(u-0.6)^2$, $g(t) = 6\sin(2\pi t)$.

在局部线性光滑中, 取 Epanechnikov 核 $K(u) = 0.75(1-u^2)_+$ 和四阶乘积核 $K_1(t,u) = K_0(t)K_0(u)$, 其中 $K_0(t) = (3/8)(3-5t^2)$, $t \in [-1,1]$. 取 $I_{\mathcal{X}}(x) = I_{[-0.5,0.5]^2}(x)$. 由 8.5 节中的交叉核实法选取带宽. 采用两种方法估计系数函数, 联系函数和参数, 即, 8.2 节所提出的逐步方法和 Wong 等 (2008) 所提的平均方法. 用 200 次重复计算得到估计曲线 $\hat{\theta}(\cdot)$ 和 $\hat{g}(\cdot)$ 以及估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$. 计算 $\hat{\theta}(\cdot)$ 和 $\hat{g}(\cdot)$ 时取样本量 n =100, 计算 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 时取样本量 n =50, 100, 150. 模拟结果见表 8.6.1 和图 8.6.1.

由表 8.6.1 可以看出本章所提出的 β 的逐步估计的偏差、标准差和均方误差均远小于 Wong 等 (2008) 的平均方法. 此外, 估计的偏差、标准差和均方误差都随着样本量的增大而减小.

图 8.6.1(a)~(c) 表示在一个典型样本下 $\theta_1(\cdot)$, $\theta_2(\cdot)$ 和 $g(\cdot)$ 的逐步和平均估计. 该典型样本由下列方式选取, 其根均方误差 (RMSE) 是 200 次计算的结果, 其中

$$\mathrm{RMSE}_j = \left[n_{\mathrm{grid}}^{-1} \sum_{k=1}^{n_{\mathrm{grid}}} [\hat{\theta}_j(u_k) - \theta_j(u_k)]^2 \right]^{1/2}, \quad j = 1, 2,$$

$$\mathrm{RMSE}g = \left[n_{\mathrm{grid}}^{-1} \sum_{k=1}^{n_{\mathrm{grid}}} [\hat{g}(u_k) - g(u_k)]^2 \right]^{1/2},$$

$\{u_k, k = 1, \cdots, n_{\mathrm{grid}}\}$ 为适当的格子点. 图 8.6.1(d) 是 200 个 RMSE 的箱线图. 图 8.6.1 表明逐步估计方法优于平均估计方法.

表 8.6.1 β_1 和 β_2 的估计的偏差 (Bias)、标准差 (SD) 和均方误差 (MSE)

n	β	逐步方法			平均方法		
		Bias	SD	MSE	Bias	SD	MSE
50	β_1	−0.0125	0.0320	0.00118	0.0126	0.0453	0.00221
	β_2	0.0063	0.0176	0.00035	0.0157	0.0495	0.00270
100	β_1	−0.0029	0.0098	0.00010	0.0092	0.0326	0.00115
	β_2	0.0016	0.0056	0.00003	0.0124	0.0368	0.00137
150	β_1	−0.0026	0.0087	0.00028	0.0068	0.0225	0.00055
	β_2	0.0015	0.0049	0.00003	0.0083	0.0248	0.00068

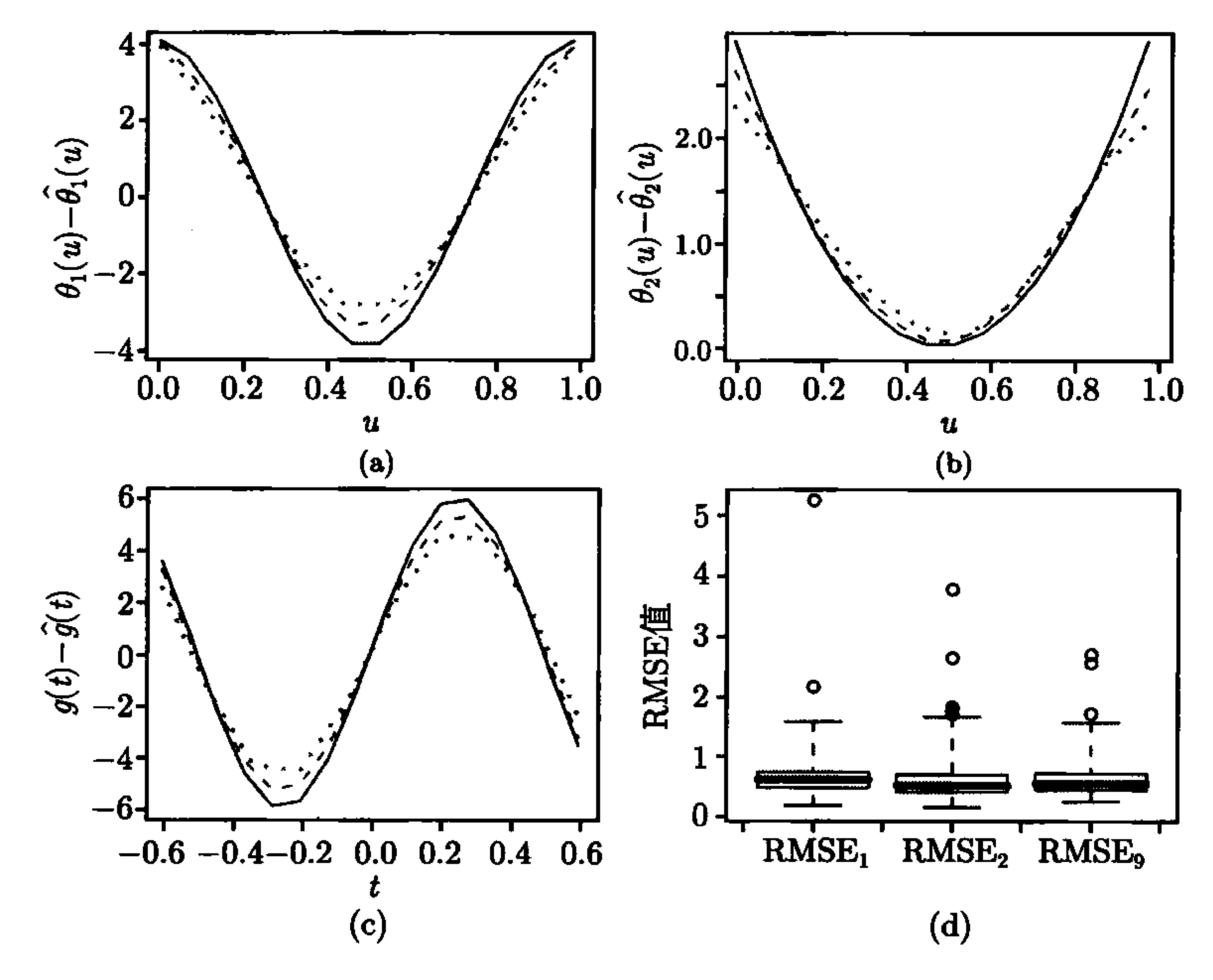


图 8.6.1 $n = 100$ 时的模拟结果. (a)~(c): 真实曲线 (实线), 逐步方法的估计曲线 (虚线), 平均方法的估计曲线 (点线). (a) 对 $\theta_1(\cdot)$, (b) 对 $\theta_2(\cdot)$, (c) 对 $g(\cdot)$; (d) $\theta_1(\cdot)$, $\theta_2(\cdot)$ 和 $g(\cdot)$ 的估计量的 RMSE 箱线图

8.6.2 实际例子

下面通过分析波士顿住房数据进一步说明我们方法的有效性. 波士顿住房数据源自 Harrison 和 Rubinfield(1978) 分析空气污染对住房价格影响的工作. 关于数据的描述已在引言中给出, 本节采用部分变系数单指标模型 (8.1.2) 来拟合该数据. 研究表明 PTRATIO 和 AGE 均影响住房价格, 这是一个合理的结果. 为计算他们的估计量, 核函数取为 8.6.1 小节中定义的 Epanchinikov 核和乘积核, 取 8.5 节中定

义的带宽, 并取 $I_{\mathcal{X}}(x)=1$. 经计算, β_1 和 β_2 的估计值分别为 -0.9980 和 -0.0631. 计算结果展示在图 8.6.2 和图 8.6.3 中.

图 8.6.2(a)~(e) 描绘了基于逐步估计方法得到的系数函数的估计曲线 (实线) 和相应的 95% 的逐点置信区间 (虚线). 从该图可以看出, 估计曲线不是很光滑, 这可能是由交叉核实方法选择的带宽过小造成的. 相关系数 $R^2=0.8206$, 残差标准差为 3.9318. 该结果表明在学校较拥挤的地段住房价格往往较低. 另外, 基准函数的值接近于 0, 因此我们可以将基准函数从模型中去掉. 图 8.6.2(f) 给出 $g(\cdot)$ 的估计曲线 (实线) 和 95% 的逐点置信区间 (虚线), 横轴为 $\hat{\beta}^{\mathrm{T}}x$ 的值, 纵轴为 $g(\cdot)$ 的估计值. 由图 8.6.2(f) 可以看出, $\hat{g}$ 的值随着指标值 —PTRATIO 和 AGE 的线性组合— 的增加而减少, 这蕴涵 PTRATIO 和 AGE 影响住房价格. 图 8.6.3 给出基于逐步估计方法的得到的 (β_1,β_2) 的 90%, 95% 和 99% 的置信域.

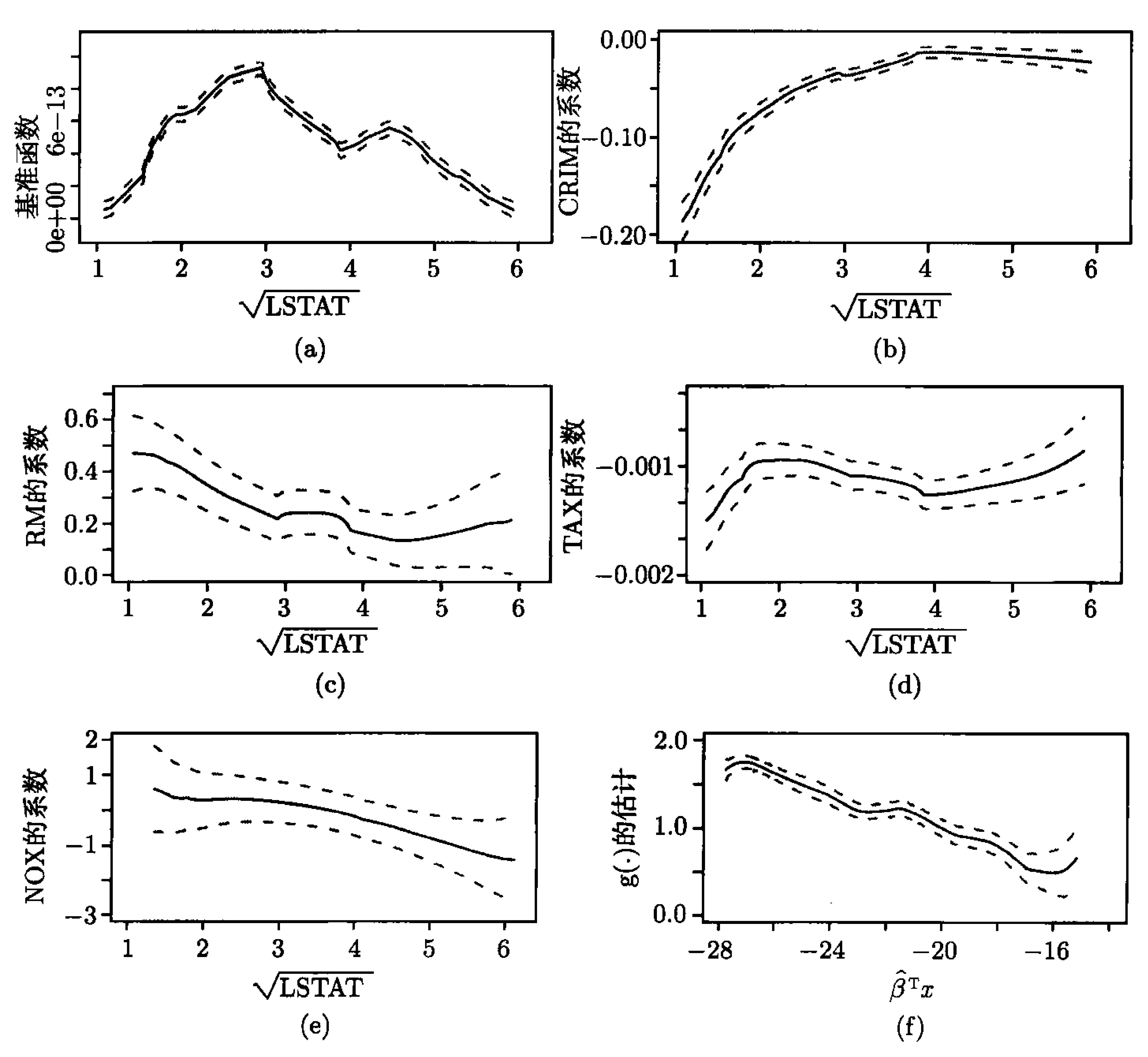


图 8.6.2　波士顿住房数据的应用. 基于逐步估计方法得到的系数函数的估计曲线 (实线) 和 95% 的逐点置信区间 (虚线). (a)~(e) 是对基准函数和系数函数; (f) 是对联系函数 $g(\cdot)$

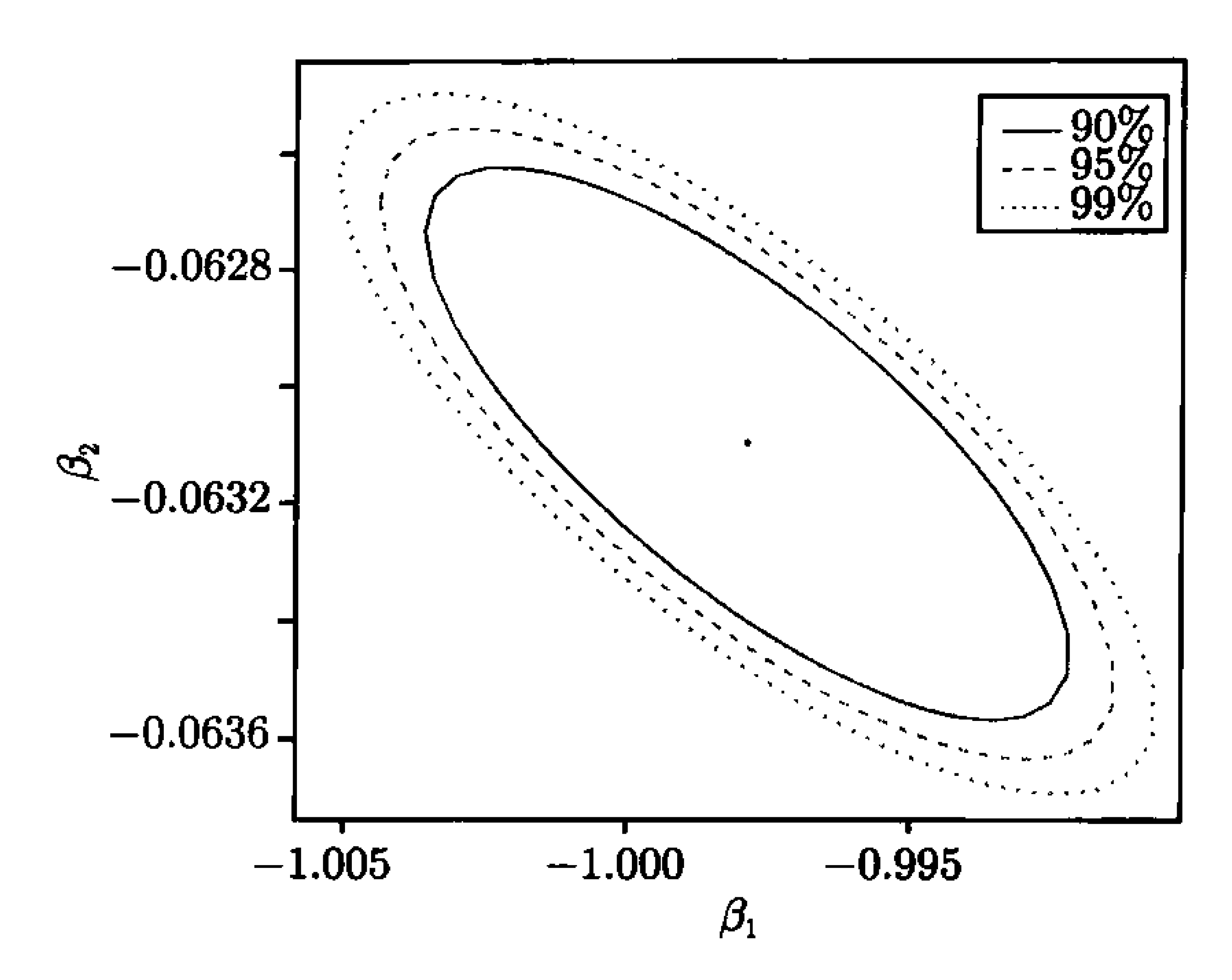


图 8.6.3 波士顿住房数据的应用. 基于逐步估计方法得到的 (β_1, β_2) 的 90%、95% 和 99% 的置信域

8.7 定理的证明

为获得定理的证明, 需要下面的引理, 这个引理可以参见 (Mack and Silverman, 1982).

引理 8.7.1 记 $(X_1, Y_1), \cdots, (X_n, Y_n)$ 为 iid 的随机向量序列, 其中 Y_i 为一维随机变量. 进一步, 假定 $E|Y_1|^s < \infty$ 和 $\sup\limits_x \int |y|^s f(x,y)\mathrm{d}y < \infty$, 其中 $f(\cdot,\cdot)$ 为 (X_1, Y_1) 的联合密度函数. 记 $K(\cdot)$ 为具有有界支撑的有界正函数, 且满足 Lipschitz 条件. 如果对 $\epsilon < 1 - s^{-1}$, 有 $n^{2\epsilon-1}h \to \infty$, $0 < h \to 0$, 则

$$\sup_x \left| \frac{1}{n} \sum_{i=1}^n \{K_h(X_i - x)Y_i - E[K_h(X_i - x)Y_i]\} \right| = O_P\left(\left[\frac{\log(1/h)}{nh} \right]^{1/2} \right).$$

定理 8.3.1 的证 可以证明, 对任意的 $\beta \in \mathcal{B}_n$ 和每一个 $j = 0, 1, 2, 3$, 有

$$R_{n,j}(u_0; \beta) = f(u_0)\Omega(u_0)\mu_j + o_P(1), \tag{8.7.1}$$

其中 $\mu_j = \int u^j K(u)\mathrm{d}u$. 事实上, 由 (8.2.5) 式 $R_{n,j}(u_0; \beta)$ 的定义可知

$$\begin{aligned} R_{n,j}(u_0;\beta) = & \frac{1}{n}\sum_{i=1}^n \left[Z_i - m_2(\beta_0^{\mathrm{T}} X_i, U_i)\right]^{\otimes 2} \left(\frac{U_i - u_0}{h_2}\right)^j K_{h_2}(U_i - u_0) \\ & + \frac{2}{n}\sum_{i=1}^n \left[Z_i - m_2(\beta_0^{\mathrm{T}} X_i, U_i)\right] \widetilde{m}_2^{\mathrm{T}}(\beta_0^{\mathrm{T}} X_i, U_i) \left(\frac{U_i - u_0}{h_2}\right)^j K_{h_2}(U_i - u_0) \\ & + \frac{1}{n}\sum_{i=1}^n \widetilde{m}_2(\beta_0^{\mathrm{T}} X_i, U_i)\widetilde{m}_2^{\mathrm{T}}(\beta_0^{\mathrm{T}} X_i, U_i) \left(\frac{U_i - u_0}{h_2}\right)^j K_{h_2}(U_i - u_0) \end{aligned}$$

$$\equiv L_1(u_0) + L_2(u_0) + L_3(u_0), \tag{8.7.2}$$

其中 $\widetilde{m}_2(\beta_0^{\mathrm{T}} X_i, U_i) = m_2(\beta_0^{\mathrm{T}} X_i, U_i) - \hat{m}_2(\beta^{\mathrm{T}} X_i, U_i; \beta)$. 易证

$$L_1(u_0) = f(u_0)\Omega(u_0)\mu_j + o_P(1). \tag{8.7.3}$$

利用 (Einmahl and Mason, 2005) 中的定理 2, 可以证明, 对 $\nu = 1, 2$, 依概率 1 有

$$\sup_{(x,\beta)\in\mathcal{X}\times\mathcal{B}_n, u\in\mathcal{N}(\beta_0)} \|\hat{m}_\nu(\beta^{\mathrm{T}} x, u; \beta) - m_\nu(\beta_0^{\mathrm{T}} X, u)\| = O\left(\left[\frac{\log n}{nh_1}\right]^{1/2} + h_1^2\right), \tag{8.7.4}$$

其中 $\mathcal{X}$ 和 $\mathcal{B}_n$ 分别在条件 (C8.3.1) 和 8.3 第一段中定义. 由 (8.7.4) 式可得 $L_2(u_0) = o_P(1)$, $L_3(u_0) = o_P(1)$. 再结合 (8.7.2) 和 (8.7.3) 式可证得 (8.7.1) 式成立. 由 (8.7.1) 式, 对任意 $\beta \in \mathcal{B}_n$ 成立

$$R_n(u_0; \beta) = R(u_0) + o_P(1), \tag{8.7.5}$$

其中 $R(u_0) = f(u_0)\Omega(u_0) \otimes \mathrm{diag}(1, \nu_2)$, $\otimes$ 表示 Kronecker 乘积.

为证明 $\check{\theta}(u_0; \beta)$ 的渐近正态性, 我们用 $\widetilde{Y}_i - \widetilde{Z}_i\theta(U_i)$ 替代 $\widetilde{Y}_i$ 来中心化 (8.2.4) 式的向量 $\eta_n(u_0; \beta)$. 记 $\eta_n^* \equiv \eta_n^*(u_0; \beta) = \left(\eta_{n,0}^{*\mathrm{T}}(u_0; \beta), \eta_{n,1}^{*\mathrm{T}}(u_0; \beta)\right)^{\mathrm{T}}$, 其中

$$\eta_{n,j}^*(u_0; \beta) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{Z}_i \left(\frac{U_i - u_0}{h_2}\right)^j K_{h_2}(U_i - u_0)\left[\widetilde{Y}_i - \theta^{\mathrm{T}}(U_i)\widetilde{Z}_i\right]. \tag{8.7.6}$$

记 $\eta_{n,j} = \eta_{n,j}(u_0; \beta)$, $\eta_n = \eta_n(u_0; \beta)$, $R_n = R_n(u_0; \beta)$, $R_{n,j} = R_{n,j}(u_0; \beta)$. 将 $\theta(U_i)$ 在 u_0 点 Taylor 展开, 由 (8.2.4), (8.7.1) 和 (8.7.6) 式可得, 对任意 $\beta \in \mathcal{B}_n$ 成立

$$\eta_{n,0} - \eta_{n,0}^* = \theta(u_0)R_{n,0} + h_2\theta'(u_0)R_{n,1} + \frac{1}{2}h_2^2\theta''(u_0)R_{n,2} + o_P(h_2^2),$$

$$\eta_{n,1} - \eta_{n,1}^* = \theta(u_0)R_{n,1} + h_2\theta'(u_0)R_{n,2} + \frac{1}{2}h_2^2\theta''(u_0)R_{n,3} + o_P(h_2^2).$$

于是

$$\eta_n - \eta_n^* = R_n\begin{pmatrix}\theta(u_0)\\ h_2\theta'(u_0)\end{pmatrix} + \frac{h_2^2}{2}\theta''(u_0)\begin{pmatrix}R_{n,2}\\ R_{n,3}\end{pmatrix} + o_P(h_2^2). \tag{8.7.7}$$

因而, 由 (8.2.3), (8.7.1), (8.7.5) 和 (8.7.7) 式可得, 对任意 $\beta \in \mathcal{B}_n$, 有

$$\begin{pmatrix}\check{\theta}(u_0; \beta) - \theta(u_0)\\ h_2[\check{\theta}'(u_0; \beta) - \theta'(u_0)]\end{pmatrix} = R^{-1}\eta_n^* + \frac{h_2^2}{2}\theta''(u_0)\begin{pmatrix}\nu_2\\ \dfrac{\mu_3}{\nu_2}\end{pmatrix} + o_P(h_2^2),$$

其中 $R = R(u_0)$. 显然

$$\breve{\theta}(u_0;\beta) - \theta(u_0) = f^{-1}(u_0)\Omega^{-1}\eta_{n,0}^*(u_0;\beta) + \frac{1}{2}h_2^2\nu_2\theta''(u_0) + o_P(h_2^2). \tag{8.7.8}$$

因此, 由 (8.7.8) 式, 为证明定理 8.3.1, 只需证明, 对任意 $\beta \in \mathcal{B}_n$ 成立

$$\sqrt{nh_2}\,\eta_{n,0}^*(u_0;\beta) \xrightarrow{D} N(0,\ \nu_0 f(u_0)\sigma^2(u_0)\Omega(u_0)), \tag{8.7.9}$$

简单计算可知

$$\begin{aligned}\eta_{n,0}^*(u_0;\beta) =& \frac{1}{n}\sum_{i=1}^n \widetilde{Z}_i\varepsilon_i K_{h_2}(U_i - u_0)\\ &+\frac{1}{n}\sum_{i=1}^n \widetilde{Z}_i\left[m_1(\beta_0^{\mathrm{T}}X_i, U_i) - \hat{m}_1(\beta^{\mathrm{T}}X_i, U_i;\beta)\right]K_{h_2}(U_i - u_0)\\ &+\frac{1}{n}\sum_{i=1}^n \widetilde{Z}_i\left[m_2(\beta_0^{\mathrm{T}}X_i, U_i) - \hat{m}_2(\beta^{\mathrm{T}}X_i, U_i;\beta)\right]\theta(U_i)K_{h_2}(U_i - u_0)\\ \equiv& J_1(u_0) + J_2(u_0) + J_3(u_0).\end{aligned} \tag{8.7.10}$$

$J_1(u_0)$ 可以分解为

$$\begin{aligned}J_1(u_0) =& \frac{1}{n}\sum_{i=1}^n \left[Z_i - m_2(\beta_0^{\mathrm{T}}X_i, U_i)\right]\varepsilon_i K_{h_2}(U_i - u_0)\\ &+\frac{1}{n}\sum_{i=1}^n \left[m_2(\beta_0^{\mathrm{T}}X_i, U_i) - \hat{m}_2(\beta_0^{\mathrm{T}}X_i, U_i;\beta_0)\right]\varepsilon_i K_{h_2}(U_i - u_0)\\ &+\frac{1}{n}\sum_{i=1}^n \left[\hat{m}_2(\beta_0^{\mathrm{T}}X_i, U_i;\beta_0) - \hat{m}_2(\beta^{\mathrm{T}}X_i, U_i;\beta)\right]\varepsilon_i K_{h_2}(U_i - u_0)\\ \equiv& J_{11}(u_0) + J_{12}(u_0) + J_{13}(u_0).\end{aligned} \tag{8.7.11}$$

直接计算可得 $E[J_{11}(u_0)] = 0$, $nh_2\mathrm{cov}[J_{11}(u_0)] = \nu_0 f(u_0)\sigma^2(u_0)\Omega(u_0) + o(1)$. 易证 $\sqrt{nh_2}J_{11}(u_0)$ 满足 Cramer-Wold 定理的条件 (Serfling, 1980, §1.5.2 中定理) 和 Lindeberg 条件 (Serfling, 1980, §1.9.2 中定理). 因此

$$\sqrt{nh_2}J_{11}(u_0) \xrightarrow{D} N(0, \nu_0 f(u_0)\sigma^2(u_0)\Omega(u_0)). \tag{8.7.12}$$

下面证明

$$E\left[\hat{m}_{2j}(\beta_0^{\mathrm{T}}X_i, U_i;\beta_0) - m_{2j}(X_i\beta_0, U_i)\right]^2 = O\left((nh_1)^{-1} + h_1^4\right), \tag{8.7.13}$$

其中 $\hat{m}_{2j}$ 和 m_{2j} 分别为 $\hat{m}_2$ 和 m_2 的第 j 个分量, $j = 1, \cdots, q$. 令 $J_{12,j}(u_0)$ 表示 $J_{12}(u_0)$ 的第 j 个分量.

事实上, 对固定的 $(t,u)\in\mathcal{T}\times\mathcal{N}(u_0)$, 令

$$\hat{r}(t,u)=\frac{1}{nh_1}\sum_{i=1}^{n}I_{\mathcal{X}}(X_i)Z_{ij}K_1\left(\frac{\beta_0^{\mathrm{T}}X_i-t}{h_1^{1/2}},\frac{U_i-u}{h_1^{1/2}}\right),$$

$$\hat{f}(t,u)=\frac{1}{nh_1}\sum_{i=1}^{n}I_{\mathcal{X}}(X_i)K_1\left(\frac{\beta_0^{\mathrm{T}}X_i-t}{h_1^{1/2}},\frac{U_i-u}{h_1^{1/2}}\right).$$

从而

$$\begin{aligned}|\hat{m}_{2j}(t,u;\beta_0)-m_{2j}(t,u)|&\leqslant\left|\hat{m}_{2j}(t,u)-\frac{E[\hat{r}(t,u)]}{E[\hat{f}(t,u)]}\right|+\left|\frac{E[\hat{r}(t,u)]}{E[\hat{f}(t,u)]}-m_{2j}(t,u)\right|\\&\leqslant\frac{1}{|\hat{f}(t,u)|}|\hat{r}(t,u)-E[\hat{r}(t,u)]|\\&\quad+\frac{|E[\hat{r}(t,u)]|}{|\hat{f}(t,u)E[\hat{f}(t,u)]|}|\hat{f}(t,u)-E[\hat{f}(t,u)]|\\&\quad+\frac{1}{|E[\hat{f}(t,u)]|}|E[\hat{r}(t,u)]-m_{2j}(t,u)f(t,u)|\\&\quad+\frac{|m_{2j}(t,u)|}{|E[\hat{f}(t,u)]|}|f(t,u)-E[\hat{f}(t,u)]|.\end{aligned}$$

易知

$$\begin{aligned}E[\hat{r}(t,u)-E[\hat{r}(t,u)]]^2&=O\left((nh_1)^{-1}\right),\\E[\hat{f}(t,u)-E[\hat{f}(t,u)]]^2&=O\left((nh_1)^{-1}\right),\\E[\hat{r}(t,u)]&=m_{2j}(t,u)f(t,u)+O(h_1^2),\\E[\hat{f}(t,u)]&=f(t,u)+O(h_1^2).\end{aligned}$$

由 (Einmahl and Mason, 2005) 中定理 1 可知, 依概率 1, 有

$$\sup_{(t,u)\in\mathcal{T}_{\beta_0}\times\mathcal{N}(u_0)}|\hat{f}(t,u))-E[\hat{f}(t,u)]|=O\left((nh_1/\log n)^{-1/2}\right).$$

因此, 由条件 (C8.3.4), 当 n 充分大时, 依概率 1, 有

$$\inf_{(t,u)\in\mathcal{T}_{\beta_0}\times\mathcal{N}(u_0)}|\hat{f}(t,u))|\geqslant|E[\hat{f}(t,u)]|-\sup_{(t,u)\in\mathcal{T}\times\mathcal{N}(u_0)}|\hat{f}(t,u))-E[\hat{f}(t,u)]|>0.$$

由以上结果及引理 8.7.1 即知 (8.7.13) 式成立.

由 (8.7.13) 式易得

$$\begin{aligned}E[J_{12,j}(u_0)]^2&\leqslant\frac{1}{n^2h_2^2}\sum_{i=1}^{n}E[m_{2j}(\beta_0^{\mathrm{T}}X_i,U_i)-\hat{m}_{2j}(\beta_0^{\mathrm{T}}X_i,U_i;\beta_0)]^2\\&=O\left((n^2h_1h_2^2)^{-1}+n^{-1}h_1^4h_2^{-2}\right).\end{aligned}$$

从而由定理 8.3.1 中的带宽条件可得

$$J_{12}(u_0) = O_P\left((nh_2)^{-1/2}\right). \tag{8.7.14}$$

注意到

$$\begin{aligned}||J_{13}(u_0)|| \leqslant & \sup_{(x,\beta)\in\mathcal{X}\times\mathcal{B}_n, u\in\mathcal{N}(u_0)} ||\hat{m}_2'(\beta^{\mathrm{T}}x, u; \beta)||||\beta-\beta_0|| \\ & \times \frac{1}{nh_2}\sum_{i=1}^{n}|\varepsilon_i|K\left(\frac{U_i-u_0}{h_2}\right) \\ = & O_P(n^{-1/2}).\end{aligned}$$

因此, 由 (8.7.11), (8.7.12), (8.7.14) 式和 Slutsky 定理可证得

$$\sqrt{nh_2}J_1(u_0) \xrightarrow{D} N(0, \nu_0 f(u_0)\sigma^2(u_0)\Omega(u_0)). \tag{8.7.15}$$

类似地, 可以证明 $\sqrt{nh_2}J_2(u_0) \xrightarrow{P} 0$, $\sqrt{nh_2}J_3(u_0) \xrightarrow{P} 0$. 再结合 (8.7.10) 和 (8.7.15) 式可证得 (8.7.9) 式成立. 这就完成了定理 8.3.1 的证明.

定理 8.3.2 的证 由引理 8.7.1, 类似于 (8.7.1) 式的证明可得, 对 $j = 0, 1, 2, 3$ 以及对 $u \in \mathcal{N}(u_0)$ 和 $\beta \in \mathcal{B}_n$ 一致有

$$R_{n,j}(u;\beta) = f(u)\Omega(u)\mu_j + O_P\left(\left[\frac{\log n}{nh_1}\right]^{1/2} + \left[\frac{\log n}{nh_2}\right]^{1/2} + h_1^2\right), \tag{8.7.16}$$

由 (8.7.16) 式立得, 对 $u \in \mathcal{N}(u_0)$ 和 $\beta \in \mathcal{B}_n$ 一致有

$$R_n(u;\beta) = R(u) + O_P\left(\left[\frac{\log n}{nh_1}\right]^{1/2} + \left[\frac{\log n}{nh_2}\right]^{1/2} + h_1^2\right).$$

由上式及事实

$$(A + h_1 B)^{-1} = A^{-1} - h_1 A^{-1} B A^{-1} + O(h_1^2),$$

可得, 对 $u \in \mathcal{N}(u_0)$ 和 $\beta \in \mathcal{B}_n$ 一致有

$$R_n^{-1}(u;\beta) = R^{-1}(u) + O_P\left(\left[\frac{\log n}{nh_1}\right]^{1/2} + \left[\frac{\log n}{nh_2}\right]^{1/2} + h_1^2\right). \tag{8.7.17}$$

由引理 8.7.1 和 (8.7.4) 式, 对 $u \in \mathcal{N}(u_0)$ 和 $\beta \in \mathcal{B}_n$ 一致有

$$\eta_n^*(u;\beta) = O_P\left(\left[\frac{\log n}{nh_1}\right]^{1/2} + \left[\frac{\log n}{nh_2}\right]^{1/2} + h_1^2\right). \tag{8.7.18}$$

由 (8.7.16)~(8.7.18) 式, 类似于 (8.7.8) 式的证明即可证得定理 8.3.2.

定理 8.3.3 的证 证明分两步: 第一步证明 β_0 的最小二乘估计 $\hat{\beta}$ 的存在性, 第二部证明 $\hat{\beta}$ 的渐近正态性.

(I) **存在性.** 我们证明如下事实: 在条件 (C8.3.1)~(C8.3.9) 下, $Q(\beta)$ 在 $\mathcal{B}_{1n}$ 内依概率 1 有唯一最小值, 其中 $Q(\beta)$ 在 (8.2.9) 式中定义, $\mathcal{B}_{1n}=\{\beta:||\beta-\beta_0||=B_1 n^{-1/2}\}$, $0<B_1<\infty$. 由 (8.2.13) 式, 只需证明在 $\mathcal{B}_{2n}$ 内存在 $\beta_0^{(r)}$ 的最小二乘估计, 其中 $\mathcal{B}_{2n}=\{\beta^{(r)}:||\beta^{(r)}-\beta_0^{(r)}||=B_2 n^{-1/2}\}$, $0<B_2<\infty$. 下面证明这一事实. 记 $M=(\theta^{\mathrm{T}}(U_1)Z_1,\cdots,\theta^{\mathrm{T}}(U_n)Z_n)^{\mathrm{T}}$, $\check{M}=(\check{\theta}^{\mathrm{T}}(U_1;\beta)Z_1,\cdots,\check{\theta}^{\mathrm{T}}(U_n;\beta)Z_n)^{\mathrm{T}}$, $Y=(Y_1,\cdots,Y_n)^{\mathrm{T}}$, $s_i(\beta)=(W_{n1}(\beta^{\mathrm{T}}X_i;\beta),\cdots,W_{nn}(\beta^{\mathrm{T}}X_i;\beta))^{\mathrm{T}}$, $S_\beta=(s_1(\beta),\cdots,s_n(\beta))^{\mathrm{T}}$, $W^*=\mathrm{diag}\{I_{\mathcal{X}}(X_1),\cdots,I_{\mathcal{X}}(X_n)\}$. 有

$$\begin{aligned}
Q(\beta)&=(Y-\check{M})^{\mathrm{T}}(I-S_\beta)^{\mathrm{T}}W^*(I-S_\beta)(Y-\check{M})\\
&=(Y-M)^{\mathrm{T}}(I-S_\beta)^{\mathrm{T}}W^*(I-S_\beta)(Y-M)\\
&\quad-2(Y-M)^{\mathrm{T}}(I-S_\beta)^{\mathrm{T}}W^*(I-S_\beta)(\check{M}-M)\\
&\quad+(\check{M}-M)^{\mathrm{T}}(I-S_\beta)^{\mathrm{T}}W^*(I-S_\beta)(\check{M}-M)\\
&\equiv Q_1(\beta)-Q_2(\beta)+Q_3(\beta).
\end{aligned}$$

由定理 8.3.2, 直接计算可得, 对 $\beta\in\mathcal{B}_{1n}$, 一致有 $Q_2(\beta)=Q_0+o_P(\sqrt{n})$ 和 $Q_3(\beta)=o_P(\sqrt{n})$, 其中 Q_0 是不依赖于 β 的常数. 这意味着对 $\beta\in\mathcal{B}_{1n}$, 一致有 $Q(\beta)=Q_1(\beta)-Q_0+o_P(\sqrt{n})$. 记 $R^*(\beta^{(r)})=(-1/2)\dfrac{\partial Q(\beta)}{\partial\beta^{(r)}}$, 则有 $R^*(\beta^{(r)})=R^{**}(\beta^{(r)})+o_P(\sqrt{n})$, 其中 $R^{**}(\beta^{(r)})=(-1/2)\dfrac{\partial Q_1(\beta)}{\partial\beta^{(r)}}$. 解方程 $R^{**}(\beta^{(r)})=0$ 可以得到 $\beta_0^{(r)}$ 的估计量 $\hat{\beta}^{(r)}$. 我们也可以通过解方程 $R(\beta^{(r)})=0$ 得到 $\hat{\beta}^{(r)}$, 其中

$$R(\beta^{(r)})=\sum_{i=1}^n I_{\mathcal{X}}(X_i)\left[Y_i-\theta^{\mathrm{T}}(U_i)Z_i-\tilde{g}(\beta^{\mathrm{T}}X_i;\beta)\right]\tilde{g}'(\beta^{\mathrm{T}}X_i;\beta)J_{\beta^{(r)}}^{\mathrm{T}}X_i,$$

$$\tilde{g}(t;\beta)=\sum_{i=1}^n W_{ni}(t;\beta)\left[Y_i-\theta^{\mathrm{T}}(U_i)Z_i\right],$$

$$\tilde{g}'(t;\beta)=\sum_{i=1}^n \widetilde{W}_{ni}(t;\beta)\left[Y_i-\theta^{\mathrm{T}}(U_i)Z_i\right].$$

可以证明

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^*}||R(\beta^{(r)})-U(\beta_0^{(r)})+nV(\beta^{(r)}-\beta_0^{(r)})||=o_P\big(\sqrt{n}\big), \tag{8.7.19}$$

其中 $\mathcal{B}_n^* = [\beta^{(r)} : ||\beta^{(r)} - \beta_0^{(r)}|| \leqslant C^* n^{-1/2}]$, C^* 为大于 0 的常数, V 在条件 (C8.3.10) 中定义, 且

$$U(\beta_0^{(r)}) = \sum_{i=1}^n I_{\mathcal{X}}(X_i)\varepsilon_i g'(\beta_0^{\mathrm{T}} X_i) J_{\beta_0^{(r)}}^{\mathrm{T}} \left[X_i - E(X_i|\beta_0^{\mathrm{T}} X_i)\right]. \tag{8.7.20}$$

事实上, 分解 $R(\beta^{(r)})$ 可得

$$\begin{aligned}
R(\beta^{(r)}) =& \sum_{i=1}^n I_{\mathcal{X}}(X_i)\varepsilon_i g'(\beta_0^{\mathrm{T}} X_i) J_{\beta^{(r)}}^{\mathrm{T}} \left[X_i - E(X_i|\beta_0^{\mathrm{T}} X_i)\right]. \\
&+ \sum_{i=1}^n I_{\mathcal{X}}(X_i)\varepsilon_i \left[\tilde{g}'(\beta^{\mathrm{T}} X_i;\beta) - g'(\beta_0^{\mathrm{T}} X_i)\right] \beta) J_{\beta^{(r)}}^{\mathrm{T}} X_i \\
&- \sum_{i=1}^n I_{\mathcal{X}}(X_i) g'(\beta_0^{\mathrm{T}} X_i) J_{\beta^{(r)}}^{\mathrm{T}} X_i \left[\tilde{g}(\beta^{\mathrm{T}} X_i;\beta) - \tilde{g}(\beta_0^{\mathrm{T}} X_i;\beta_0)\right] \\
&- \sum_{i=1}^n I_{\mathcal{X}}(X_i) g'(\beta_0^{\mathrm{T}} X_i) J_{\beta^{(r)}}^{\mathrm{T}} \left\{X_i \left[\tilde{g}(\beta_0^{\mathrm{T}} X_i;\beta_0) - g(\beta_0^{\mathrm{T}} X_i)\right] - \varepsilon_i \mu_1(\beta_0^{\mathrm{T}} X_i)\right\} \\
&- \sum_{i=1}^n I_{\mathcal{X}}(X_i) \left[\tilde{g}(\beta^{\mathrm{T}} X_i;\beta) - g(\beta_0^{\mathrm{T}} X_i)\right] \left[\tilde{g}'(\beta^{\mathrm{T}} X_i;\beta) - g'(\beta_0^{\mathrm{T}} X_i)\right] J_{\beta^{(r)}}^{\mathrm{T}} X_i \\
\equiv& R_1(\beta^{(r)}) + R_2(\beta^{(r)}) - R_3(\beta^{(r)}) - R_4(\beta^{(r)}) - R_5(\beta^{(r)}).
\end{aligned} \tag{8.7.21}$$

注意到 $J_{\beta^{(r)}} - J_{\beta_0^{(r)}} = O_P(n^{-1/2})$ 对所有 $\beta^{(r)} \in \mathcal{B}_n^*$ 成立, 于是有

$$\sup_{\beta^{(r)} \in \mathcal{B}_n^*} ||R_1(\beta^{(r)}) - U(\beta_0^{(r)})|| = o_P(\sqrt{n}), \tag{8.7.22}$$

其中 $\mathcal{B}_n^*$ 和 $U(\beta_0^{(r)})$ 分别在 (8.7.19) 和 (8.7.20) 式中定义.

下面考虑 $R_2(\beta^{(r)})$. 简单计算可得

$$\begin{aligned}
R_2(\beta^{(r)}) =& \sum_{i=1}^n I_{\mathcal{X}}(X_i)\varepsilon_i \left[\tilde{g}'(\beta^{\mathrm{T}} X_i;\beta) - \tilde{g}'(\beta_0^{\mathrm{T}} X_i;\beta_0)\right] J_{\beta^{(r)}}^{\mathrm{T}} X_i \\
&+ \sum_{i=1}^n I_{\mathcal{X}}(X_i)\varepsilon_i \left[\tilde{g}'(\beta_0^{\mathrm{T}} X_i;\beta_0) - g'(\beta_0^{\mathrm{T}} X_i)\right] J_{\beta^{(r)}}^{\mathrm{T}} X_i \\
\equiv& R_{21}(\beta^{(r)}) + R_{22}(\beta_0^{(r)}).
\end{aligned}$$

因为对所有 $\beta^{(r)} \in \mathcal{B}_n^*$ 和 $\beta \in \mathcal{B}_n$, $||\beta^{(r)} - \beta_0^{(r)}|| \leqslant C^* n^{-1/2}$ 蕴涵 $||\beta - \beta_0|| \leqslant C^* n^{-1/2}$. 又由 (8.2.13) 和 (8.2.14) 式易知

$$\hat{\beta} - \beta_0 = J_{\beta_0^{(r)}}(\hat{\beta}^{(r)} - \beta_0^{(r)}) + O_P(n^{-1}). \tag{8.7.23}$$

于是对 $\beta^{(r)} \in \mathcal{B}_n^*$, 一致有

$$\begin{aligned}R_{21}(\beta^{(r)}) =& \sum_{i=1}^{n} I_{\mathcal{X}}(X_i)\varepsilon_i \tilde{g}''(X_i^{\mathrm{T}}\bar{\beta};\bar{\beta}) J_{\beta^{(r)}}^{\mathrm{T}} X_i [X_i - E(X_i|\beta_0^{\mathrm{T}}X_i)]^{\mathrm{T}}(\beta-\beta_0)\\ =& o_P(\sqrt{n}).\end{aligned}$$

又由引理 3.10.2 和引理 3.10.3 可得

$$E\bigg[\sum_{j=1}^{n}\widetilde{W}_{nj}(\beta_0^{\mathrm{T}}X_i;\beta_0)g(\beta_0^{\mathrm{T}}X_i) - g'(\beta_0^{\mathrm{T}}X_i)\bigg]^2 = O(h_3^2),$$

$$\sum_{j=1}^{n} E\left[\widetilde{W}_{nj}^2(\beta_0^{\mathrm{T}}X_i;\beta_0)\right] = O\left((nh_3^3)^{-1}\right).$$

从而

$$\begin{aligned}& E\left[n^{-1}R_{22}^2(\beta_0^{(r)})\right]\\ \leqslant & cn^{-1}\sum_{i=1}^{n} E\left[\sum_{j=1}^{n}\widetilde{W}_{nj}(\beta_0^{\mathrm{T}}X_i;\beta_0)g(\beta_0^{\mathrm{T}}X_j) - g'(\beta_0^{\mathrm{T}}X_i)\right]^2\\ & + cn^{-1}\sum_i\sum_j\sum_k\sum_l E\left[\widetilde{W}_{nj}(\beta_0^{\mathrm{T}}X_i;\beta_0)\widetilde{W}_{nl}(\beta_0^{\mathrm{T}}X_k;\beta_0)\varepsilon_i\varepsilon_k\varepsilon_j\varepsilon_l\right]\\ \leqslant & ch_3^2 + cn^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[\widetilde{W}_{nj}^2(\beta_0^{\mathrm{T}}X_i;\beta_0)\right]\\ \leqslant & ch_3^2 + c(nh_3^3)^{-1} \longrightarrow 0.\end{aligned}$$

因此, $R_{22}(\beta_0^{(r)}) = o_P(\sqrt{n})$. 这就证明了

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^*} ||R_2(\beta^{(r)})|| = o_P(\sqrt{n}). \tag{8.7.24}$$

对于 $R_3(\beta^{(r)})$, 由 $\beta^{(r)} - \beta_0^{(r)}$ 的 Taylor 展式及一个适当的均值 $\bar{\beta}^{(r)} \in \mathcal{B}_n^*$, $\bar{\beta} = \bar{\beta}(\bar{\beta}^{(r)})$, 有

$$\begin{aligned}R_3(\beta^{(r)}) =& \sum_{i=1}^{n} I_{\mathcal{X}}(X_i)g'(\beta_0^{\mathrm{T}}X_i)\tilde{g}'(\bar{\beta}^{\mathrm{T}}X_i;\bar{\beta})J_{\beta^{(r)}}^{\mathrm{T}}X_i[X_i - E(X_i|\beta_0^{\mathrm{T}}X_i)]^{\mathrm{T}}J_{\bar{\beta}^{(r)}}\\ & \times\left(\beta^{(r)} - \beta_0^{(r)}\right)\\ =& \sum_{i=1}^{n} I_{\mathcal{X}}(X_i)g'(\beta_0^{\mathrm{T}}X_i)\left[\tilde{g}'(\bar{\beta}^{\mathrm{T}}X_i;\bar{\beta}) - g'(\beta_0^{\mathrm{T}}X_i)\right]\end{aligned}$$

$$
\begin{aligned}
&\times J_{\beta^{(r)}}^{\mathrm{T}} X_i[X_i - E(X_i|\beta_0^{\mathrm{T}} X_i)]^{\mathrm{T}} J_{\bar{\beta}^{(r)}}\left(\beta^{(r)} - \beta_0^{(r)}\right) \\
&+ \sum_{i=1}^{n} g'(\beta_0^{\mathrm{T}} X_i)^2 J_{\beta^{(r)}}^{\mathrm{T}} X_i[X_i - E(X_i|\beta_0^{\mathrm{T}} X_i)]^{\mathrm{T}} J_{\bar{\beta}^{(r)}}\left(\beta^{(r)} - \beta_0^{(r)}\right) \\
\equiv& R_{31}(\beta^{(r)}, \bar{\beta}^{(r)}) + R_{32}(\beta^{(r)}, \bar{\beta}^{(r)}).
\end{aligned}
$$

类似于定理 8.3.2 的证明可得

$$
\sup_{(x,\beta)\in\mathcal{X}\times\mathcal{B}_n} \left|\tilde{g}'(\beta^{\mathrm{T}} x;\beta) - g'(\beta_0^{\mathrm{T}} X)\right| = o_P(1).
$$

由上式可得

$$
\sup_{\beta^{(r)},\bar{\beta}^{(r)}\in\mathcal{B}_n^*} ||R_{31}(\beta^{(r)}, \bar{\beta}^{(r)})|| = o_P\left(\sqrt{n}\right),
$$

$$
\sup_{\beta^{(r)},\bar{\beta}^{(r)}\in\mathcal{B}_n^*} ||R_{32}(\beta^{(r)}, \bar{\beta}^{(r)}) - nV\left(\beta^{(r)} - \beta_0^{(r)}\right)|| = o_P\left(\sqrt{n}\right).
$$

因此

$$
\sup_{\beta^{(r)}\in\mathcal{B}_n^*} ||R_3(\beta^{(r)}) - nV\left(\beta^{(r)} - \beta_0^{(r)}\right)|| = o_P\left(\sqrt{n}\right). \tag{8.7.25}
$$

现在考虑 $R_4(\beta^{(r)})$. 记 $R_4(\beta^{(r)}) = J_{\beta^{(r)}}^{\mathrm{T}} R_4^*(\beta^{(r)})$. 令 $R_{4,s}^*$ 表示 $R_4^*(\beta^{(r)})$ 的第 s 个分量. 首先, 由引理 3.10.2 和引理 3.10.3 可知

$$
E\left[\sum_{j=1}^{n} W_{nj}(\beta_0^{\mathrm{T}} X_i;\beta_0) g(\beta_0^{\mathrm{T}} X_j) - g(\beta_0^{\mathrm{T}} X_i)\right]^2 = O\left(h_3^4\right),
$$

$$
E\left[\sum_{j=1}^{n} W_{ni}(\beta_0^{\mathrm{T}} X_j;\beta_0) \phi(\beta_0^{\mathrm{T}} X_j) - \phi(\beta_0^{\mathrm{T}} X_i)\right]^2 = O\left(h_3^{1/2}\right),
$$

$$
\sum_{j=1}^{n} E\left[W_{nj}^2(\beta_0^{\mathrm{T}} X_i;\beta_0)\right] = O\left((nh_3)^{-1}\right),
$$

其中 $\phi(\cdot) = g'(\cdot)\mu_s(\cdot)$, $\mu_s(\cdot)$ 是 $\mu(\cdot)$ 的第 s 个分量, $W_{nj}(\cdot;\cdot)$ 在 (8.2.7) 式中定义. 因此

$$
\begin{aligned}
n^{-1} E\left(R_{4,s}^{*2}\right) \leqslant& cn^{-1} \sum_{i=1}^{n} E\left[\sum_{j=1}^{n} W_{ni}(\beta_0^{\mathrm{T}} X_j;\beta_0) g'(\beta_0^{\mathrm{T}} X_j) X_{js} - \phi(\beta_0^{\mathrm{T}} X_i)\right]^2 \\
&+ c\sum_{i=1}^{n} E\left[\sum_{j=1}^{n} W_{nj}(\beta_0^{\mathrm{T}} X_i;\beta_0) g(\beta_0^{\mathrm{T}} X_j) - g(\beta_0^{\mathrm{T}} X_i)\right]^2
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant cn^{-1}\sum_{i=1}^{n}E\left[\sum_{j=1}^{n}W_{ni}(\beta_0^{\mathrm{T}}X_j;\beta_0)g'(\beta_0^{\mathrm{T}}X_j)\left[X_{js}-\mu_s(\beta_0^{\mathrm{T}}X_j)\right]\right]^2\\
&\quad+cn^{-1}\sum_{i=1}^{n}E\left[\sum_{j=1}^{n}W_{ni}(\beta_0^{\mathrm{T}}X_j;\beta_0)\phi(\beta_0^{\mathrm{T}}X_j)-\phi(\beta_0^{\mathrm{T}}X_i)\right]^2+cnh_3^4\\
&\leqslant c(nh_3)^{-1}+ch_3^{1/2}+cnh_3^4\to 0.
\end{aligned}
$$

于是

$$
\sup_{\beta^{(r)}\in\mathcal{B}_n^*}||R_4(\beta^{(r)})||=o_P(\sqrt{n}), \tag{8.7.26}
$$

类似可得

$$
\sup_{\beta^{(r)}\in\mathcal{B}_n^*}||R_5(\beta^{(r)})||=o_P(\sqrt{n}). \tag{8.7.27}
$$

结合 (8.7.21)~(8.7.27) 式可知 (8.7.19) 式成立.

对任意一个 $\beta_*^{(r)}\in\mathcal{B}_{2n}$ 及 $\mathcal{B}_{2n}$ 中有待确定的常数 B_2, 由 (8.7.19) 式可得

$$
\begin{aligned}
&(\beta_*^{(r)}-\beta_0^{(r)})^{\mathrm{T}}R^*(\beta_*^{(r)})\\
&=(\beta_*^{(r)}-\beta_0^{(r)})^{\mathrm{T}}U(\beta_0^{(r)})-n(\beta_*^{(r)}-\beta_0^{(r)})^{\mathrm{T}}V(\beta_*^{(r)}-\beta_0^{(r)})+o_P(1).
\end{aligned} \tag{8.7.28}
$$

下面的讨论类似于 Weisberg 和 Welsh(1994), 需要反复使用 Ortega 和 Rheinboldt(1973) 中的 (6.3.4) 式. 注意到 (8.7.27) 式中的主要项由 B_2^2 决定, 因为 $\sqrt{n}||\beta_*^{(r)}-\beta_0^{(r)}||=B_2$, 而 $|(\beta_*^{(r)}-\beta_0^{(r)})^{\mathrm{T}}U(\beta_0^{(r)})|=B_2O_P(1)$, $n(\beta_*^{(r)}-\beta_0^{(r)})^{\mathrm{T}}V(\beta_*^{(r)}-\beta_0^{(r)})\sim B_2^2$. 因此, 对任何给定的 $\delta>0$, 若 B_2 的取值足够大, 则依概率 $1-\delta$ 有 $(\beta_*^{(r)}-\beta_0^{(r)})^{\mathrm{T}}R^*(\beta_0^{(r)})<0$. 由 δ 的任意性, 并类似于 Welsh(1989) 中定理 5.1 的证明, 可以证明 $\beta_0^{(r)}$ 在 $\mathcal{B}_{2n}$ 中的最小二乘估计的存在性. 这里略去其证明细节.

(II) **渐近正态性.** 由 (I) 可知 $\hat\beta^{(r)}$ 是方程 $R^*(\beta^{(r)})=0$ 在 $\mathcal{B}_{2n}$ 内的一个解, 即 $R^*(\hat\beta^{(r)})=0$. 由 (8.7.19) 式可得

$$
0=U(\beta_0^{(r)})-nV(\hat\beta^{(r)}-\beta_0^{(r)})+o_P(\sqrt{n}\ ),
$$

所以

$$
\sqrt{n}(\hat\beta^{(r)}-\beta_0^{(r)})=V^{-1}n^{-1/2}U(\beta_0^{(r)})+o_P(1).
$$

$$
\sqrt{n}(\hat\beta-\beta_0)=J_{\beta_0^{(r)}}V^{-1}n^{-1/2}U(\beta_0^{(r)})+o_P(1).
$$

结合 (8.7.20) 式, 由中心极限定理和 Slutsky 定理即可得定理 8.3.3.

推论 8.3.3 的证 由定理 8.3.3 可知 $\|\hat{\beta}-\beta_0\|=O_P\big(n^{-1/2}\big)$ 及

$$|\hat{\beta}^{\mathrm{T}}\beta_0-1|=|(\hat{\beta}-\beta_0)^{\mathrm{T}}\beta_0|\leqslant\|\hat{\beta}-\beta_0\|=O_P\big(n^{-1/2}\big).$$

这就完成了推论 8.3.3 的证明.

定理 8.3.4 的证 记

$$\hat{\mu}_{2j}(t;\beta)=\sum_{i=1}^{n}W_{ni}(t;\beta)Z_{ij}.$$

类似于 (8.7.4) 式和定理 8.3.2 的证明可得

$$\sup_{(x,\beta)\in\mathcal{X}\times\mathcal{B}_n}\left|\hat{\mu}_{2j}(\beta^{\mathrm{T}}x;\beta)-\mu_{2j}(\beta^{\mathrm{T}}x)\right|=o_P(1),$$

$$\sup_{(x,\beta)\in\mathcal{X}\times\mathcal{B}_n}\left|\tilde{g}(\beta^{\mathrm{T}}x;\beta)-g(\beta^{\mathrm{T}}x)\right|=O_P\left(\left[\frac{\log n}{nh_3}\right]^{1/2}+h_3^2\right),$$

其中 $\mathcal{X}$ 和 $\mathcal{B}_n$ 分别在条件 (C8.3.1) 和 8.3 节第一段中定义. 推论 8.3.2 说明

$$\begin{aligned}
&\sup_{(x,\beta)\in\mathcal{X}\times\mathcal{B}_n}\left|\hat{g}(\beta^{\mathrm{T}}x)-g(\beta^{\mathrm{T}}x)\right|\\
\leqslant&\sum_{j=1}^{q}\sup_{(x,\beta)\in\mathcal{X}\times\mathcal{B}_n}\left|\hat{\mu}_{2j}(\beta^{\mathrm{T}}x;\beta)\right|\sup_{u\in\mathcal{N}(u_0)}\left|\hat{\theta}_j(u)-\theta_j(u)\right|\\
&+q\sup_{(x,\beta)\in\mathcal{X}\times\mathcal{B}_n}\left|\tilde{g}(\beta^{\mathrm{T}}x;\beta)-g(\beta^{\mathrm{T}}x)\right|\\
=&O_P\left(\left[\frac{\log n}{nh_3}\right]^{1/2}+h_3^2+n^{-2/5}(\log n)^{1/2}\right),
\end{aligned}$$

从而定理 8.3.4 得证.

定理 8.3.5 的证 记 $\varphi_i=[\theta(U_i)-\hat{\theta}(U_i)]^{\mathrm{T}}Z_i$, $M_i=g(\beta_0^{\mathrm{T}}X_i)-\hat{g}(\hat{\beta}^{\mathrm{T}}X_i)$. 由 (8.2.11) 中 $\hat{\sigma}^2(u)$ 的定义可知

$$\begin{aligned}
\hat{\sigma}^2(u)=&\hat{f}^{-1}(u)n^{-1}\sum_{i=1}^{n}(\varphi_i+M_i+\varepsilon_i)^2K_{h_2}(U_i-u)\\
=&\hat{f}^{-1}(u)n^{-1}\sum_{i=1}^{n}(\varphi_i+M_i)^2K_{h_2}(U_i-u)\\
&+2\hat{f}^{-1}(u)n^{-1}\sum_{i=1}^{n}\varepsilon_i(\varphi_i+M_i)K_{h_2}(U_i-u)\\
&+\hat{f}^{-1}(u)n^{-1}\sum_{i=1}^{n}\varepsilon_i^2K_{h_2}(U_i-u)\\
\equiv&I_1(u)+I_2(u)+I_3(u).
\end{aligned}$$

由定理 8.3.2, 引理 8.7.1 及 (8.7.4) 式可证

$$\sup_{u\in\mathcal{N}(u_0)}|I_1(u)| = O_P\left(\frac{\log n}{nh_3} + h_3^4 + n^{-4/5}\log n\right),$$

$$\sup_{u\in\mathcal{N}(u_0)}|I_2(u)| = O_P\left(\left[\frac{\log n}{nh_3}\right]^{1/2} + h_3^2 + n^{-2/5}(\log n)^{1/2}\right),$$

$$\sup_{u\in\mathcal{N}(u_0)}|I_3(u) - \sigma^2(u)| = O_P\left(n^{-2/5}\right).$$

综上所述, 我们得到了定理 8.3.5 中 $\hat{\sigma}^2(u)$ 的一致收敛速度. 在定理 8.3.3 的假设条件下, 依概率有 $I_3(u_0) \to \sigma^2(u_0)$, 以及 $I_\nu(u_0) = o_P(1)$, $\nu = 2,3$. 因此, $\hat{\sigma}^2(u_0)$ 是 $\sigma^2(u_0)$ 的相合估计.

参 考 文 献

Arnold S F. 1981. The Theory of Linear Models and Multivariate Analysis. New York: John Wiley & Sons

Ahmad I, Leelahanon S, Li Q. 2005. Efficient estimation of a semiparametric partially linear varying coefficient model. The Annals of Statistics, 33: 258～283

Cai Z W, Fan J, Li R Z. 2000. Efficient estimation and inferences for varying-coefficient models . Journal of the American Statistical Association, 95: 888～902

Cai Z W, Fan J, Yao Q. 2000. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association, 95: 941～956

Carroll R J, Fan J, Gijbels I, et al.. 1997. Generalized partially linear single index models. Journal of the American Statistical Association, 92: 477～489

Chiou J M, Müller H G. 1998. Quasi-likelihood regression with unknown link and variance functions. Journal of the American Statistical Association, 93: 1376～1387

Chiang C T, Rice J A, Wu C O. 2001. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. Journal of the American Statistical Association, 93: 961～994

Einmahl U, Mason D M. 2005. Uniform in bandwidth consistency of kernel-type function estimators. The Annals of Statistics, 33: 1380～1403

Fan J, Gijbels I. 1996. Local Polynomial Modeling and Its Applications. London: Chapman and Hall

Fan J, Huang T. 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli, 11: 1031～1057

Fan J, Zhang C, Zhang J. 2001. Generalized likelihood ratio statistics and Wilks Phenomenon. The Annals of Statistics, 29: 153～193

Fan J, Zhang W Y. 2000. Simultaneous confidence bands and hypothesis testing in varying-coefficient models. Scandinavian Journal of statistics, 27: 715～731

Green P J, Silverman B W. 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. London: Chapman and Hall

Härdle W, Hall P, Ichimura H. 1993. Optimal smoothing in single-index models. The Annals of Statistics, 21: 157～178

Härdle W, Stoker T M. 1989. Investing smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association, 84: 986～995

Harrison D, Rubinfeld D L. 1978. Hedonic housing pries and the demand for clean air. Journal of Environmental Economics and Management, 5: 81～102

Hastie T J, Tibshirani R. 1993. Varying-coefficient models. Journal of the Royal Statistical Society, Series B, 55: 757～796

Huang J Z, Wu C O, Zhou L. 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika, 89: 111～128

Huang Z S. 2011. Empirical likelihood-based inference in varying-coefficient single index models. Journal of the Korean Statistical Society, 40(2): 205～215

Hoover D R, Rice J A, Wu C O, et al.. 1998. Nonparametric smoothing estimates of time varying coefficient models with longitudinal data. Biometrika, 85: 809～822

Li K C. 1991. Sliced inverse regression for dimension reduction. Journal of the American Statistical Association, 86: 316～342

Mack Y P, Silverman B W. 1982. Weak and strong uniform consistency of kernel regression estimates. Z. Wahrsch. verw. Gebiete, 61: 405～415

Ortega J M, Rheinboldt W C. 1973. Iterative Solution of Nonlinear Equations in Several Variables. New York: Academic Press

Serfling R. 1980. Approximation Theorems of Mathematical Statistics. New York: Wiley

Wahba G. 1984. Partial spline models for semiparametric estimation of functions of several variables. In *Statistical Analysis of Time Series*, Proceedings of the Japan U.S. Joint Seminar, Tokyo, 319～329. Tokyo: Institute of Statistical Mathematics

Wang J L, Xue L G, Zhu L X, et al.. 2010. Estimation for a partial-linear single-index model. The Annals of Statistics, 38: 246～272

Wang Q H, Xue L G. 2011. Statistical inference partially-varying-coefficient single-index model. Journal of Multivariate Analysis, 102(1): 1～19

Weisberg S, Welsh A H. 1994. Adapting for the Missing Linear Link. The Annals of Statistics, 22: 1674～1700

Welsh A H. 1989. On M-processes and M-estimation. The Annals of Statistics, 17: 337～361. (Correction, 1990, 18: 1500)

Wong H, Ip W, Zhang R Q. 2008. Varying-coefficient single-index model. Computational Statistics and Data Analysis, 52: 1458～1476

Wu C O, Chiang C T, Hoover D R. 1998. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. Journal of the American Statistical Association, 93: 1388~1402

Xia Y C, Li W K. 1999. On single-index coefficient regression models. Journal of the American Statistical Association, 94: 1275~1285

Xia Y C, Li W K, Tong H. 2007. Estimating threshold variables using nonparametric methods. Statistica Sinica, 17: 265~287

Xia Y C, Zhang W Y, Tong H. 2004. Efficient estimation for semivarying-coefficient models. Biometrika, 91: 661~681

Xue L G, Zhu L X. 2006. Empirical likelihood for single-index model. Journal of Multivariate Analysis, 97: 1295~1312

Xue L G, Zhu L X. 2007. Empirical Likelihood for a varying coefficient model with longitudinal data. Journal of the American Statistical Association, 102: 642~654

Yu Y, Ruppert D. 2002. Penalized spline estimation for partially linear single-index models. Journal of the American Statistical Association, 97: 1042~1054

Zhu L X, Fang K T. 1996. Asymptotics for the kernel estimates of sliced inverse regression. The Annals of Statistics, 24: 1053-1067

Zhu L X, Xue L G. 2006. Empirical likelihood confidence regions in a partially linear single-index model. Journal of the Royal Statistical Society, Series B, 68: 549~570

第 9 章　单指标混合效应模型

9.1 引　　言

混合效应模型广泛应用于分析相关数据, 如纵向数据和重复测量数据等. 一些作者研究了线性和非线性混合效应模型, 有关文献可参看 (Harville, 1977; Lindstrom and Bates, 1990; Ke and Wang, 2001) 等. 对截面数据, Gu 和 Ma(2005) 提出了非参数混合效应模型

$$Y = \eta(X) + Z^{\mathrm{T}}b + \varepsilon,$$

其中回归函数 $\eta(x)$ 是一般域 $\mathcal{X}$ 上的光滑函数. 使用惩罚 (非加权) 最小二乘方法估计函数项 $\eta(x)$, 而方差矩阵和光滑样条中的调整参数 λ 都被看作调整参数而不加估计. 然而, 当 X 的维数很高时, 将会出现 "维数灾祸" 问题. 为了弥补这两点不足, 我们考虑单指标混合效应模型

$$Y_{ij} = g(\beta_0^{\mathrm{T}} X_{ij}) + Z_{ij}^{\mathrm{T}} b_i + \varepsilon_{ij}, \quad i = 1, \cdots, n, j = 1, \cdots, m, \tag{9.1.1}$$

其中 β_0 是 $p \times 1$ 指标系数向量, b_i 是均值为 0 且协方差阵为 D 的独立 $q \times 1$ 随机效应向量, $g(\cdot)$ 是未知联系函数, 诸 ε_{ij} 是具有均值 0 和方差 $\sigma_\varepsilon^2 > 0$ 的独立随机误差. 这里 D 是依赖于参数向量 ϕ 的正定矩阵, X_{ij} 和 Y_{ij} 是可观测的随机变量, Z_{ij} 为固定设计矩阵. 假设 b_i 和 ε_{ij} 相互独立. 为了使模型可识别, 通常假设 $||\beta_0|| = 1$ 且 β_0 的第一个非 0 分量是正数, 其中 $||\cdot||$ 表示 Euclid 模. 本章将阐述同时估计参数 β_0、函数 $g(\cdot)$ 以及 D 中的方差分量和 σ_ε^2 的一般问题.

单指标模型在多元非参回归研究中是一个重要工具, 它通过寻找多元协变量 X 的单变量指标来捕捉高维数据的特征而避免了 "维数灾祸". 单指标模型已经应用于许多领域, 如计量经济学中的离散选择分析和生物学中的剂量反应模型 (Härdle et al., 1993). 在截面数据中, 许多作者研究了单指标模型的统计推断问题, 并得到了一些令人满意的结果, 部分文献包括 (Li, 1991; Härdle et al., 1993; Ichimura, 1993; Xia and Li, 1999; Naik and Tsai, 2000; Hristache et al., 2001; Xia et al., 2002; Xia, 2006; Xue and Zhu, 2006; Cui et al., 2011) 等. 本章使用带随机效应的单指标模型来分析实际问题中常常遇到的带有随机效应的数据.

一些现有文献研究了广义线性、非参数和半参数混合效应模型的统计推断问题. 主要文献包括 (Zeger and Diggle, 1994; Ruckstuhl et al., 2000; Ke and Wang,

2001; Cai et al., 2002; Wu and Zhang, 2002; Hall and Maiti, 2006; Jiang, 2007) 等. 然而, 关于单指标混合效应模型的研究并不多见. Honorá和 Kyriazidou(2000) 与 Carro(2007) 针对动态面板数据离散选择模型提出了一些估计方法. Bai 等 (2009) 研究了纵向数据的单指标模型, 他们基于惩罚样条和二次推断函数提出了指标系数和联系函数的估计方法. Liang 和 Zeger(1986) 针对广义线性模型提出了广义估计方程 (GEEs) 方法, 并在时间相依的较弱条件下得到了指标系数及其方差的相合估计. 在不需要指定个体观测的联合分布的情形下可以得到 GEEs, 从而避开了使用多元正态分布的得分方程.

本章将 GEEs 的思想应用于单指标混合效应模型中. 为估计指标系数 β_0, 我们提出了一个新的估计方程, 它考虑了边界效应和约束条件 $\|\beta_0\|=1$. 在对样本联合分布的较弱假定下, 得到了 β_0 的 $\sqrt{n}$ 相合估计. 同时也构造了 D 中方差分量和 σ_ε^2 的 $\sqrt{n}$ 相合估计. 并得到了 $g(\cdot)$ 的估计量的渐近正态性和一致收敛速度. 本章的主要内容取材于文献 (Pang and Xue, 2012).

9.2 估计方法

9.2.1 参数和非参数分量的估计

本节首先关注指标系数 β_0 的估计. 设 $Y_i=(Y_{i1},\cdots,Y_{im})^{\mathrm{T}}$, $X_i=(X_{i1},\cdots,X_{im})^{\mathrm{T}}$, $Z_i=(Z_{i1},\cdots,Z_{im})^{\mathrm{T}}$, $G(X_i\beta_0)=(g(X_{i1}^{\mathrm{T}}\beta_0),\cdots,g(X_{im}^{\mathrm{T}}\beta_0))^{\mathrm{T}}$, $\varepsilon_i=(\varepsilon_{i1},\cdots,\varepsilon_{im})^{\mathrm{T}}$. 那么, 模型 (9.1.1) 可表示为

$$Y_i=G(X_i\beta_0)+Z_ib_i+\varepsilon_i,\quad i=1,\cdots,n.$$

该模型表明了诸 Y_i 是互相独立的, $E(Y_i|X_i)=G(X_i\beta_0)$ 且 $\mathrm{cov}(Y_i|X_i)=Z_iDZ_i^{\mathrm{T}}+\sigma_\varepsilon^2I_m$, 其中 I_m 是 $m\times m$ 单位矩阵. 估计 β_0 的一个自然的想法是在 $\|\beta\|=1$ 下关于 β 极小化目标函数

$$R_n(\beta)=\frac{1}{n}\sum_{i=1}^{n}\{Y_i-G(X_i\beta)\}^{\mathrm{T}}W(X_i\beta)V_i^{-1}\{Y_i-G(X_i\beta)\},\tag{9.2.1}$$

其中 $V_i=Z_iDZ_i^{\mathrm{T}}+\sigma_\varepsilon^2I_m$ 是 $Z_ib_i+\varepsilon_i$ 的协方差阵, $W(X_i\beta)=\mathrm{diag}\{w(X_{i1}\beta),\cdots,w(X_{im}\beta)\}$, $w(\cdot)$ 是具有紧支撑 $\mathcal{U}_w$ 的有界权函数, 它是用来控制在估计 $g(\cdot)$ 和 $g'(\cdot)$ 时的边界效应. 为简单方便起见, 设 $w(\cdot)$ 为 $\mathcal{U}_w$ 上的示性函数.

然而, 在 (9.2.1) 式中, 联系函数 $g(\cdot)$ 和 D 中的方差分量 ϕ 以及 σ_ε^2 都是未知的, 我们不能直接得到 β_0 的估计. 下一节中将证明可以以 $\sqrt{n}$ 的速度估计 σ_ε^2 和 D 中的方差分量 ϕ. 因此, 在建立和分析 β_0 和 $g(\cdot)$ 的估计量时将他们看作已知量. 然后提出一个平行的方法来迭代地估计 β_0 和 $g(\cdot)$.

如果 $g(\cdot)$ 是已知的, 那么 (9.2.1) 式是一个约束的最小二乘问题. 由于约束条件 $||\beta_0|| = 1$, 因此函数 $g(\beta^{\mathrm{T}}X_{ij})$ 在点 β_0 处不存在导数. 使用 3.10 节 "去掉一分量" 方法可以将约束的最小二乘转换为 $\mathbf{R}^{p-1}$ 上的无约束的最小二乘问题, 但需要注意的是 3.10 节是处理独立数据, 方差 σ^2 作为讨厌参数并不出现在极小化问题中. 相比较而言, 本章模型中 D 和 σ_ε^2 都不能作为讨厌参数, 必须给出它们的估计量.

为清楚起见, 这里对记号作详细说明. 不失一般性, 本章假设真实参数 β_0 的第 r 个分量是正的. 设 $\beta^{(r)} = (\beta_1, \cdots, \beta_{r-1}, \beta_{r+1}, \cdots, \beta_p)^{\mathrm{T}}$, 那么 β 关于 $\beta^{(r)}$ 的 Jacobian 矩阵可写为

$$J_{\beta^{(r)}} = \frac{\partial \beta}{\partial \beta^{(r)}} = (\gamma_1, \cdots, \gamma_p)^{\mathrm{T}},$$

其中 $\gamma_k\,(1 \leqslant k \leqslant p, k \neq r)$ 是 $p-1$ 维的单位向量, 其第 k 个分量为 1, $\gamma_r = -(1 - ||\beta^{(r)}||^2)^{-1/2}\beta^{(r)}$. 设 $X_{ij} = (X_{ij1}, \cdots, X_{ijp})^{\mathrm{T}}$, $X_{ij}^{(r)} = (X_{ij1}, \cdots, X_{ij(r-1)}, X_{ij(r+1)}, \cdots, X_{ijp})^{\mathrm{T}}$. 那么 $\beta^{\mathrm{T}}X_{ij} = X_{ij}^{(r)\mathrm{T}}\beta^{(r)} + (1 - ||\beta^{(r)}||^2)^{1/2}X_{ijr}$ 是 $\beta^{(r)}$ 的函数. 当 $g(\cdot)$ 是已知时, 解关于 $\beta^{(r)}$ 的方程组

$$Q_n(G, \beta^{(r)}) \equiv \frac{1}{n}\sum_{i=1}^{n} J_{\beta^{(r)}}^{\mathrm{T}} X_i^{\mathrm{T}} G'_\Delta(X_i\beta) W(X_i\beta) V_i^{-1}\{Y_i - G(X_i\beta)\} = 0, \quad (9.2.2)$$

可以得到 $\beta_0^{(r)}$ 的估计量, 其中 $G'_\Delta(X_i\beta) = \mathrm{diag}\{g'(X_{i1}\beta), \cdots, g'(X_{im}\beta)\}$. 一个迭代的加权最小二乘算法可以用来解上述方程组. 给定一个现行估计 $\tilde{\beta}_0^{(r)}$, 计算

$$\tilde{\beta}^{(r)} = \tilde{\beta}_0^{(r)} + B_n^{-1}(G, \tilde{\beta}_0^{(r)})Q_n(G, \tilde{\beta}_0^{(r)}),$$

其中

$$B_n(G, \beta^{(r)}) \equiv \frac{1}{n}\sum_{i=1}^{n} J_{\beta^{(r)}}^{\mathrm{T}} X_i^{\mathrm{T}} G'^2_\Delta(X_i\beta) W(X_i\beta) V_i^{-1} X_i J_{\beta^{(r)}}. \quad (9.2.3)$$

这个迭代的加权最小二乘算法求解方程 (9.2.2), 且等同于求解这个估计方程的 Newton-Raphson 算法的得分形式的 Fisher 方法. 利用 $||\beta_0|| = 1$ 和 $||\beta|| = 1$, 可以证明

$$\beta - \beta_0 = J_{\beta_0^{(r)}}(\beta^{(r)} - \beta_0^{(r)}) + O_P(n^{-1}).$$

因此, 当 $g(\cdot)$ 已知时可以得到估计 β_0 的迭代公式, 即

$$\hat{\beta}^* = \tilde{\beta} + J_{\tilde{\beta}^{(r)}} B_n^{-1}(G, \tilde{\beta}^{(r)})Q_n(G, \tilde{\beta}^{(r)}),$$

其中模为 1 的初始值 β_0 可以通过拟合线性模型而得到. 然后, 重置 $\tilde{\beta} = \hat{\beta}^*/||\hat{\beta}^*||$, 反复迭代直到收敛.

关于未知的联系函数 $g(\cdot)$, 在给定 β_0 的初始估计下可以容易计算 $g(\cdot)$ 和 $g'(\cdot)$ 的非参数估计 $\hat{g}(\cdot)$ 和 $\hat{g}'(\cdot)$. 局部线性光滑方法 (Fan and Gijbels, 1996) 能同时提供 $g(\cdot)$ 及其导数 g' 的估计量, 这里使用此方法. 选定实数集 $\mathbf{R}$ 上的核函数 $K(\cdot)$ 和趋近与 0 的带宽序列 $h = h_n$, 记 $K_h(\cdot) = h^{-1}K(\cdot/h)$. 对给定的 β, 未知函数 $g(u)$ 和 $g'(u)$ 的局部线性估计量分别定义为 $\hat{g}(u;\beta) = \hat{d}_0$ 和 $\hat{g}'(u;\beta) = \hat{d}_1$, 其中 $\hat{d}_0$ 和 $\hat{d}_1$ 为极小化下列加权平方和的解:

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\{Y_{ij} - d_0 - d_1(\beta^{\mathrm{T}}X_{ij} - u)\}^2 K_h(\beta^{\mathrm{T}}X_{ij} - u). \tag{9.2.4}$$

$\hat{d}_0$ 和 $\hat{d}_1$ 的表达式类似于 (7.2.5) 式中 Z 为一维的情形.

在已有的文献中, 称估计量 $\hat{g}$ 为 pooled 估计量, 如 (Lin and Carroll, 2000; Ruckstuhl et al., 2000; Xue, 2010). 正如这些论文中指出的那样, 在这种情况下简单的 pooled 估计量事实上与考虑相关结构的估计量一样渐近有效.

众所周知, 如果使用相同带宽, 那么 $g'(u)$ 的估计量比 $g(u)$ 的估计量的收敛速度低. 这就引起 β_0 的估计量慢于 $\sqrt{n}$ 的收敛速度. 这就需要引入另一个带宽 h_1 来控制 $g'(u)$ 的估计量的变异性. 也就是说, 下文在 $\hat{g}'(u;\beta)$ 中使用 h_1 来代替 h, 并定义 $\hat{g}'_{h_1}(u;\beta)$ 作为 $g'(u)$ 的另一个估计量.

在估计出 $\hat{g}$ 和 $\hat{g}'_{h_1}$ 之后, 我们可以使用下述得分算法来估计 β_0, 即

$$\hat{\beta} = \tilde{\beta} + J_{\tilde{\beta}^{(r)}} B_n^{-1}(\hat{G}, \tilde{\beta}^{(r)}) Q_n(\hat{G}, \tilde{\beta}^{(r)}), \tag{9.2.5}$$

并计算 $\hat{\beta}/||\hat{\beta}||$, 其值仍记作 $\hat{\beta}$, 其中用 $\hat{G}$, $\hat{G}'_{\Delta}$, $\tilde{\beta}$ 和 $\tilde{\beta}^{(r)}$ 替换 (9.2.2) 和 (9.2.3) 式中 G, G'_{Δ}, β 和 $\beta^{(r)}$ 而得到 $Q_n(\hat{G}, \tilde{\beta}^{(r)})$ 和 $B_n(\hat{G}, \tilde{\beta}^{(r)})$.

然后, 我们在 β 和 $g(\cdot)$ 之间交替计算直到满足确定的要求. 利用 $\hat{\beta}$ 可以定义 $g(\cdot)$ 的最终估计量 $\hat{g}^*(u) = \hat{g}(u;\hat{\beta})$.

9.2.2　方差分量的估计

兴趣参数的估计量的渐近方差依赖于方差分量, 因此, 需要得到方差分量的相合估计. 假设模型 (9.1.1) 的协方差阵为

$$V = \sigma_b^2 J_m J_m^{\mathrm{T}} + \sigma_\varepsilon^2 I_m, \tag{9.2.6}$$

其中 $J_m = (1, \cdots, 1)^{\mathrm{T}}$ 是关于 1 的 m 维向量. 称该模型为方差分量模型, 它在纵向数据分析中起着重要作用. 估计 σ_b^2 和 σ_ε^2 的一个有用方法是假定残差的均值为 0, 且具有与 $g(\cdot)$ 已知时相同协方差阵. 如果假设随机效应项 b_i 和误差项 ε_{ij} 具有正态分布, 观测值 Y_i 具有独立的 $N(G(X_i\beta_0), V)$ 分布. 将 $g(\cdot)$ 和 β_0 分别用它们的

估计量 $\hat{g}(\cdot)$ 和 $\hat{\beta}$ 代替; 可以写出 σ_ε^2 和 σ_b^2 的正态似然函数

$$-n(m-1)\log(\sigma_\varepsilon^2) - n\log(\sigma_\varepsilon^2 + m\sigma_b^2) - \frac{m}{\sigma_\varepsilon^2 + m\sigma_b^2}\sum_{i=1}^{n}(\bar{Y}_i - \bar{g}_i)^2$$
$$-\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{Y_{ij} - \hat{g}(X_{ij}^{\mathrm{T}}\beta;\hat{\beta}) - (\bar{Y}_i - \bar{g}_i)\right\}^2,$$

其中 $\bar{Y}_i = m^{-1}\sum\limits_{j=1}^{m} Y_{ij}$, $\bar{g}_i = m^{-1}\sum\limits_{j=1}^{m}\hat{g}(X_{ij}^{\mathrm{T}}\hat{\beta};\hat{\beta})$. 当 $\hat{\sigma}_b^2 > 0$ 时, 该似然函数在下列点处取到最大值:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{Y_{ij} - \hat{g}(X_{ij}^{\mathrm{T}}\hat{\beta};\hat{\beta}) - (\bar{Y}_i - \bar{g}_i)\right\}^2, \tag{9.2.7}$$

$$\hat{\sigma}_b^2 = \frac{1}{n}\sum_{i=1}^{n}(\bar{Y}_i - \bar{g}_i)^2 - \hat{\sigma}_\varepsilon^2/m, \tag{9.2.8}$$

而在 $\hat{\sigma}_b^2 = 0$ 时,

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{Y_{ij} - \hat{g}(X_{ij}^{\mathrm{T}}\hat{\beta};\hat{\beta})\right\}^2. \tag{9.2.9}$$

该 “似然” 方法对一般的单指标混合模型模型 (9.1.1) 同样有效. 另一个构造相同收敛速度的估计量的方法是放弃 “似然” 而采用矩方法来得到估计量. 也可以调整由于估计 $g(\cdot)$ 造成的自由度的丢失, 并得到关于 D 和 σ_ε^2 的估计量. 具体细节可参考 (Ruckstuhl et al., 2000). 可以证明, 用矩方法得到的估计量犹如 $g(\cdot)$ 和 β_0 已知时的收敛速度.

9.3 主 要 结 果

现在研究非参数分量 $g(\cdot)$、参数分量 β_0 以及 σ_ε^2 和 D 的方差分量的估计量的渐近性质. 我们首先列出下列正则条件.

(C9.3.1) 对任何 i, $(X_{i1}^{\mathrm{T}}\beta,\cdots,X_{im}^{\mathrm{T}}\beta)^{\mathrm{T}}$ 的联合密度存在; 对任何 $j_1 \neq j_2$, $\beta^{\mathrm{T}}X_{ij}$ 的边际密度 $f_j(u)$ 和 $(X_{ij_1}^{\mathrm{T}}\beta, X_{ij_2}^{\mathrm{T}}\beta)$ 的联合密度 $f_{j_1j_2}(u,s)$ 分别在 $u_0 \in \mathcal{U}_w$ 和 $(u_0,s_0) \in \mathcal{U}_w \times \mathcal{U}_w$ 处是连续可微的; 存在某个 j 使得 $f_j(u)$ 在 $u \in \mathcal{U}_w$ 和接近 β_0 的 β 点上一致有界的远离 0, 其中 $\mathcal{U}_w$ 是 $w(u)$ 的支撑集.

(C9.3.2) $g(u)$ 的二阶导数有界连续, $g_{2r}(u)$ 在 $\mathcal{U}_w$ 上满足 1 阶 Lipschitz 条件, 其中 $g_{2r}(u)$ 是 $g_2(u)$ 的第 r 个分量, 且 $g_2(u) = E(X_{ij}|\beta_0^{\mathrm{T}}X_{ij} = u)$, $1 \leqslant r \leqslant p$.

(C9.3.3) 核函数 $K(\cdot)$ 是具有有界支撑的有界对称的概率密度函数, 并满足 1 阶 Lipschitz 条件, $\int u^2K(u)\mathrm{d}u \neq 0$.

(C9.3.4) 存在 $r = \max\{4, s\}$ 使得 $E(||X_{ij}||^r) < \infty$, $E(||b_i||^r) < \infty$ 和 $E(|\varepsilon_{ij}|^r) < \infty$, 且对某个 $\epsilon < 2 - s^{-1}$ 使得 $n^{2\epsilon-1}h \to \infty$, $i = 1, \cdots, j = 1, \cdots, m$.

(C9.3.5) $nh^2/\log^2 n \to \infty$, $nh^4 \log n \to 0$; $nhh_1^3/\log^2 n \to \infty$, $nh_1^5 = O(1)$.

对正数 c_1, 令 $\mathcal{B}_n = \{\beta : ||\beta|| = 1, ||\beta - \beta_0|| \leqslant c_1 n^{-1/2}\}$. 我们的定义主要基于如下的事实, 由于希望 $\hat{\beta}$ 是 $\sqrt{n}$ 相合的, 我们将寻找方程 $Q_n(\hat{G}, \beta^{(r)}) = 0$ 的解, 它包含在 $\beta^{(r)}$ 距 $\beta_0^{(r)}$ 的阶为 $n^{-1/2}$ 的范围内. 文献 (Härdle et al., 1993; Xia and Li, 1999) 中也有类似的约束条件. 记 $\mu_l = \int u^l K(u)\mathrm{d}u$, $\nu_l = \int K^l(u)\mathrm{d}u$, $l = 1, 2$.

下述定理给出了 8.2 节中提出的估计量的渐近性质. 首先给出 $\hat{g}$ 和 $\hat{g}'$ 的一致收敛速度.

定理 9.3.1　设条件 (C9.3.1)~(C9.3.4) 成立, 则

$$\sup_{u\in\mathcal{U}_w, \beta\in\mathcal{B}_n} \left|\hat{g}(u;\beta) - g(u)\right| = O_P\left((nh/\log n)^{-1/2} + h^2\right),$$

$$\sup_{u\in\mathcal{U}_w, \beta\in\mathcal{B}_n} \left|\hat{g}'(u;\beta) - g'(u)\right| = O_P\left((nh^3/\log n)^{-1/2} + h\right).$$

下述定理 9.3.2 和定理 9.3.3 表明了估计量 $\hat{g}$ 和 $\hat{\beta}$ 的渐近正态性.

定理 9.3.2　设条件 (C9.3.1)~(C9.3.4) 成立. 如果 $nh^5 = O(1)$ 和 $\sigma^2 = \mathrm{var}(Y_{ij}|X_{ij}) > 0$, 则对任意的 $u \in \mathcal{U}_w$ 和 $\beta \in \mathcal{B}_n$, 有

$$\sqrt{nh}[\hat{g}(u;\beta) - g(u) - b(u)] \xrightarrow{D} N(0, \gamma^2(u)),$$

其中 $b(u) = (1/2)h^2\mu_2 g''(u)$, $\gamma^2(u) = \nu_2\sigma^2 / \sum_{j=1}^{m} f_j(u)$.

如果进一步假设 $nh^5 \to 0$, 则有

$$\sqrt{nh}[\hat{g}(u;\beta) - g(u)] \xrightarrow{D} N(0, \gamma^2(u)).$$

注 9.3.1　对协方差模型 (9.2.6), 我们有 $\sigma^2 = \sigma_b^2 + \sigma_\varepsilon^2 > 0$.

定理 9.3.3　设条件 (C9.3.1)~(C9.3.5) 成立, 则

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N\left(0, J_{\beta_0^{(r)}} A^- J_{\beta_0^{(r)}}^{\mathrm{T}}\right),$$

其中 A^- 表示矩阵 A 的广义逆,

$$A = E\left[J_{\beta_0^{(r)}}^{\mathrm{T}} \{X_1 - G_1(X_1\beta_0)\}^{\mathrm{T}} G_\Delta'^2(X_1\beta_0) W(X_1\beta_0) V^{-1} \{X_1 - G_1(X_1\beta_0)\} J_{\beta_0^{(r)}}\right],$$

$$G_1(X_1\beta_0)=(g_1(X_{11}\beta_0),\cdots,g_1(X_{1m}\beta_0))^{\mathrm{T}},$$

$$g_1(u)=E(X_{ij}|\beta_0^{\mathrm{T}}X_{ij}=u),\ V^{-1}=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}V_i^{-1}.$$

由定理 9.3.2 和定理 9.3.3, 可以得到下列推论 9.3.1.

推论 9.3.1 设条件 (C9.3.1)~(C9.3.5) 成立, 则对任意 $u\in\mathcal{U}_w$,

$$\sqrt{nh}[\hat{g}^*(u)-g(u)]\xrightarrow{D}N(0,\gamma^2(u)),$$

其中 $\gamma^2(u)$ 在定理 9.3.2 中定义.

由定理 9.3.3, 我们得到关于 $\hat{\beta}$ 和 β_0 的夹角的渐近结果, 它可以用来研究充分降维问题.

推论 9.3.2 设定理条件 9.3.3 成立, 则

$$|\hat{\beta}^{\mathrm{T}}\beta_0|-1=O_P(n^{-1/2}),$$

其中 $|\hat{\beta}^{\mathrm{T}}\beta_0|$ 为 $\hat{\beta}$ 与 β_0 的绝对内积.

由于模型 (9.2.6) 在纵向数据分析中的重要性, 我们给出关于 σ_ε^2 和 σ_b^2 的估计量的收敛速度. 对一般模型 (9.1.1) 也可以得到类似的结论.

定理 9.3.4 设条件 (C9.3.1)~(C9.3.5) 成立. 如果 X_{ij} 的分布具有紧支撑集, 则

$$\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2=O_P\left(n^{-1/2}\right),\tag{9.3.1}$$

$$\hat{\sigma}_b^2-\sigma_b^2=O_P\left(n^{-1/2}\right).\tag{9.3.2}$$

为了给 β_0 构造置信域, 我们需要定义 $\hat{\beta}$ 的渐近方差的估计量. A 的估计量 $\hat{A}$ 可定义为

$$\hat{A}=\frac{1}{n}\sum_{i=1}^{n}J_{\hat{\beta}^{(r)}}^{\mathrm{T}}\{X_i-\hat{G}_1(X_i\hat{\beta};\hat{\beta})\}^{\mathrm{T}}\hat{G}_\Delta'^2(X_i\hat{\beta};\hat{\beta})W(X_i\hat{\beta})\hat{V}_i^{-1}\{X_i-\hat{G}_1(X_i\hat{\beta};\hat{\beta})\}J_{\hat{\beta}^{(r)}},$$

其中 $\hat{V}_i$ 是 V_i 的估计量, $\hat{G}_1(X_i\hat{\beta};\hat{\beta})=(\hat{g}_1(\hat{\beta}^{\mathrm{T}}X_{i1};\hat{\beta}),\cdots,\hat{g}_1(\hat{\beta}^{\mathrm{T}}X_{im}\hat{\beta};\hat{\beta}))^{\mathrm{T}}$,

$$\hat{g}_1(u;\hat{\beta})=\sum_{i=1}^{n}\sum_{j=1}^{m}W_{nij}(u;\hat{\beta})X_{ij}$$

是 $g_1(u)=E(X_{ij}|\beta_0^{\mathrm{T}}X_{ij}=u)$ 的估计量. 此处

$$W_{nij}(u;\hat{\beta})=\frac{n^{-1}K_h(X_{ij}^{\mathrm{T}}\hat{\beta}-u)\{S_{n,2}(u;\hat{\beta})-[(X_{ij}^{\mathrm{T}}\hat{\beta}-u)/h]S_{n,1}(u;\hat{\beta})\}}{S_{n,0}(u;\hat{\beta})S_{n,2}(u;\hat{\beta})-S_{n,1}^2(u;\hat{\beta})},$$

其中

$$S_{n,l}(u;\beta_0)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{\beta_0^{\mathrm{T}}X_{ij}-u}{h}\right)^l K_h(\beta_0^{\mathrm{T}}X_{ij}-u),\quad l=0,1,2.$$

容易证明 $J_{\hat\beta^{(r)}}\xrightarrow{P}J_{\beta_0^{(r)}}$ 和 $\hat A\xrightarrow{P}A$. 则对任意 $p\times l$ 的 $l<p$ 满秩矩阵 H, 定理 9.3.3 表明

$$\left(n^{-1}H^{\mathrm{T}}J_{\hat\beta^{(r)}}\hat A^{-}J_{\hat\beta^{(r)}}^{\mathrm{T}}H\right)^{-1/2}H^{\mathrm{T}}(\hat\beta-\beta_0)\xrightarrow{D}N(0,I_l).$$

利用 (Arnold, 1981) 中的定理 10.2d 可以得到下列极限分布.

定理 9.3.5 设定理 9.3.4 的条件成立, 则

$$(\hat\beta-\beta_0)^{\mathrm{T}}H\left(n^{-1}H^{\mathrm{T}}J_{\hat\beta^{(r)}}\hat A^{-}J_{\hat\beta^{(r)}}^{\mathrm{T}}H\right)^{-}H^{\mathrm{T}}(\hat\beta-\beta_0)\xrightarrow{D}\chi_l^2.$$

定理 9.3.5 可以用来构造 β_0 的大样本置信域.

利用推论 9.3.1, 可以构造 $g(u_0)$ 在给定点 $u_0\in\mathcal{U}_w$ 的逐点置信区间. 这里, 我们考虑方差结构 (9.2.6), 其中 $\sigma^2=\sigma_b^2+\sigma_\varepsilon^2$. 记 $\hat g(u_0)$ 的渐近偏差和协方差依赖于 σ_ε^2, σ_b^2 和 $f_j(u_0)$. σ_ε^2 和 σ_b^2 的估计量已经由 (9.2.7)~(9.2.9) 式给出. $f_j(u_0)(j=1,\cdots,m)$ 的估计量定义为

$$\hat f_j(u_0)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{X_{ij}-u_0}{h}\right).$$

因此, 我们可以用相合估计量 $\hat f_j(u_0)$, $\hat\sigma_\varepsilon^2$ 和 $\hat\sigma_b^2$ 代替 $f_j(u_0)$, σ_ε^2 和 σ_b^2 而得到 $\hat\gamma^2(u_0)$. 因此, $\hat\gamma^2(u_0)$ 是 $\gamma^2(u_0)$ 的相合估计量. 由推论 9.3.1 可得

$$\sqrt{nh}\{\hat g(u_0;\hat\beta)-g(u_0)\}/\hat\gamma(u_0)\xrightarrow{D}N(0,1)$$

利用上述结论, 我们可以得到 $g(u_0)$ 的近似 $1-\alpha$ 置信区间.

9.4 模拟与应用

本节首先通过模拟来考察估计方法的执行情况. 然后利用单指标混合效应模型和估计方法分析实际数据. 本节在模拟和应用中使用 Epanechnikov 核.

9.4.1 模拟结果

本节使用广义交叉验证 (GCV) 方法选择估计 $g(\cdot)$ 时的带宽. GCV 统计量由下式给出:

$$\mathrm{GCV}(h)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{[Y_{ij}-\hat g_h^*(\beta^{\mathrm{T}}X_{ij})]^2}{[n^{-1}\mathrm{tr}(I-S_h)]^2},\tag{9.4.1}$$

其中 $\hat{g}_h^*(\cdot)$ 是 $g(\cdot)$ 的具有带宽 h 的估计量, S_h 是相应于带宽 h 的光滑矩阵. 极小化 GCV(h) 得到 $\hat{g}^*(\cdot)$ 的 GCV 带宽. 可以证明: 对于常数 $c>0$, $h_{\text{opt}}=cn^{-1/5}$. 因此, 我们可以使用 h_{opt} 作为 $\hat{g}^*(\cdot)$ 的最优带宽. 当计算估计量 $\hat{\beta}$ 时, 我们选择带宽,

$$h=h_{\text{opt}}n^{-1/20}(\log n)^{-1/2},\quad h_1=h_{\text{opt}}. \tag{9.4.2}$$

因为对常数 $c>0$, 有 $h=cn^{-1/4}(\log n)^{-1/2}$ 和 $h_1=cn^{-1/5}$. 该带宽满足条件 (C9.4.5). 这个带宽的选择思想可以从文献 (Carroll et al., 1997) 中找到.

例 9.4.1(低维情形) 考虑二次模型

$$Y_{ij}=12(\beta_0^{\mathrm{T}}X_{ij}-0.5)^2+1+b_i+\varepsilon_{ij},$$

其中 $\boldsymbol{\beta}_0=(\sqrt{3}/2,0.5)^{\mathrm{T}}$, X_{ij} 是二维随机向量, 其分量为 [0,1] 上独立的均匀分布, b_i 是标准正态分布随机变量, ε_{ij} 是均值为 0、方差为 0.25 的正态分布的随机变量. 此处 $g(u)=12(u-0.5)^2+1$. 个体数目 n 取为 30, 50, 100 和 200, 每个个体重复观测次数 m 设定为 3.

在这个模拟例子中, 取权函数为 $w(u)=I_{[0,(1+\sqrt{3})/2]}(u)$, 且通过拟合线性模型得到 β_0 的初始值. 本节基于 $g(\cdot)$ 和 $g'(\cdot)$ 的估计, 考虑 β_0 的估计量. 通过 500 次运行得到模拟的偏差 (Bias)、标准差 (SD) 和根均方误差 (RMSE). 模拟结果由表 9.4.1 给出.

表 9.4.1 例 9.4.1 的模拟结果. 在相关观察和忽略相关结构下, β_1 和 β_2 的估计的偏差、标准差和根均方误差

参数	n	相关观察			忽略相关		
		Bias	SD	RMSE	Bias	SD	RMSE
β_1	30	0.00318	0.03133	0.03137	0.00355	0.04061	0.04065
	50	0.00307	0.02437	0.02461	0.00289	0.02914	0.02926
	100	0.00271	0.01942	0.01959	0.00252	0.02124	0.02137
	200	0.00233	0.01287	0.01307	0.00209	0.01464	0.01477
β_2	30	−0.00909	0.06143	0.06196	−0.01154	0.07314	0.07398
	50	−0.00898	0.04602	0.04684	−0.00865	0.05222	0.05288
	100	−0.00637	0.03545	0.03598	−0.00625	0.03730	0.03779
	200	−0.00476	0.02310	0.02356	−0.00451	0.02559	0.02596

由表 9.4.1 可以看出, β_1 和 β_2 的基于相关结构的估计比忽略相关性具有较小的 SD 和 RMSE; 也就是说, 如果忽略了相关性且盲目使用独立数据的单指标模型, 就会导致较大的 SD 和 RMSE. 对小样本量 n, 尽管 SD 和 RMSE 比较大, 但改进更加显著. 我们也发现偏差、SD 和 RMSE 随 n 的增加而减小.

当 $n=100$ 时, 图 9.4.1(a) 和 (b) 分别给出了 β_1 和 β_2 的 500 个估计的直方图. 图 9.4.1(c) 和 (d) 分别给出了 β_1 和 β_2 的 500 个估计值的 Q–Q 图.

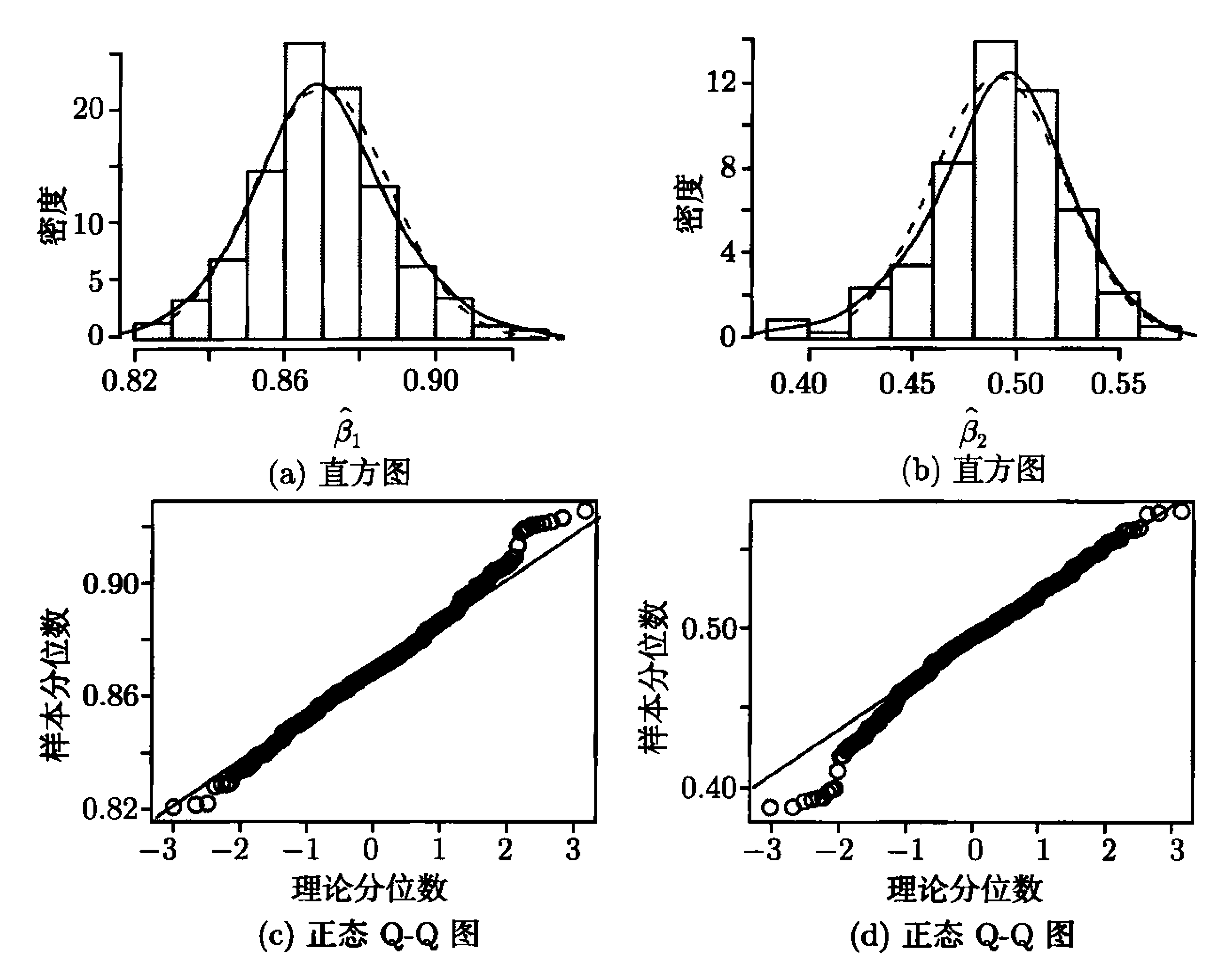


图 9.4.1 例 9.4.1 中 $n = 100$ 时的模拟结果. (a) 对 $\hat{\beta}_1$ 且 (b) 对 $\hat{\beta}_2$: 每个参数的 500 个估计的直方图, 估计的密度曲线 (实线) 和正态密度曲线 (虚线); (c) 对 $\hat{\beta}_1$ 且 (d) 对 $\hat{\beta}_2$: 每个参数的 500 个估计的 Q–Q 图

图 9.4.1 表明了估计量的渐近正态性. β_1 和 β_2 的估计量的均值分别为 0.8700 和 0.4917, 它们的 RMSE 分别为 0.0196 和 0.0360. σ_b^2 和 σ_ε^2 的估计分别为 0.8798 和 0.2801.

本节也考虑了 $g(u)$ 的估计的效果. 图 9.4.2(a) 给出了 $n = 100$ 时的真实的联系函数曲线、估计的联系函数曲线和近似 0.95 逐点置信区间. $\hat{g}^*(\cdot)$ 的估计量利用下列 RMSE 来评价:

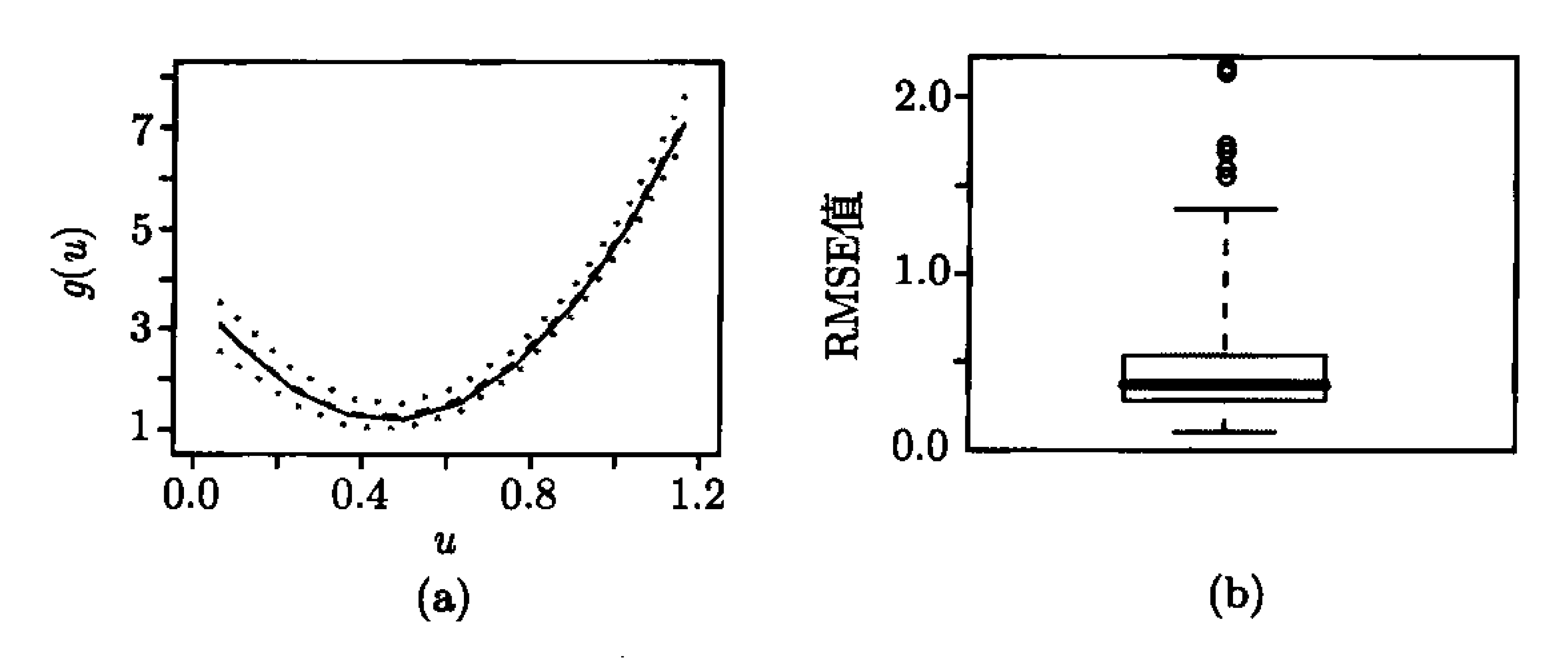


图 9.4.2 例 9.4.2 中 $n = 100$ 时的模拟结果. (a) $g(\cdot)$ 的真实曲线 (实线)、估计的曲线 (虚线) 和近似 0.95 逐点置信区间 (点线); (b) 估计 $\hat{g}(\cdot)$ 的 500 个 RMSE 的箱线图

$$\text{RMSE} = \left[n_{\text{grid}}^{-1} \sum_{k=1}^{n_{\text{grid}}} \{\hat{g}^*(u_k) - g(u_k)\}^2 \right]^{1/2},$$

其中格子点数目 n_{grid} 是 10, $\{u_k, k = 1, \cdots, n_{\text{grid}}\}$ 为等距格子点. 图 9.4.2(b) 给出了 $n = 100$ 时的 500 个 RMSE 的箱线图.

由图 9.4.2(a) 可以发现, 估计的曲线接近于真实的联系函数曲线, 且置信区间的宽度非常窄. 图 9.4.2(b) 表明联系函数的 RMSE 也很小.

例 9.4.2(高维情形) 考虑模拟模型

$$Y_{ij} = 16\exp\{-(\beta_0^{\text{T}} X_{ij})^2\} + b_i + \varepsilon_{ij},$$

其中 $\beta_0 = (0.2, 0.4, 0.6, 0.6, \sqrt{2}/5, 0, 0, 0, 0, 0)^{\text{T}}$, $p = 10$, X_{ij} 是 10 维向量, 各分量相互独立且服从 [0,1] 上的均匀分布, b_i 和 ε_{ij} 都服从均值为 0、方差分别为 $\sigma_b^2 > 0$ 和 $\sigma_\varepsilon^2 > 0$ 的正态分布, $g(u) = 16\exp(-u^2)$. 个体数目 n 分别取为 30, 50, 100 和 150, 每个个体重复观测的次数 m 为 3.

在这个模拟例子中, 取权函数为 $w(u) = I_{[0,(9+\sqrt{2})/5]}(u)$, 使用绝对内积 $|\hat{\beta}^{\text{T}}\beta_0|$ 来度量估计量 $\hat{\beta}$ 的精度. 取 $n = 30, 50, 100, 150$, $\sigma_b = 0.6, 1.2, 2$ 和 $\sigma_\varepsilon = 0.5, 1, 1.5$, 对各种样本量和噪声的组合, 做 200 次运行, 其模拟结果放在表 9.4.2 和图 9.4.3 中.

从表 9.4.2 可以看出, 内积的均值非常接近于 1, 且他们的标准差非常小. 内积随着方差分量的增加而减小, 且内积的精度随样本量的增加而增加. 图 9.4.3 表明了内积值的范围也很小, 尽管 σ_b 和 σ_ε 取值比较大.

表 9.4.2 例 9.4.2 的模拟结果. 估计值 $\hat{\beta}$ 和真实值 β_0 的内积的均值和标准差 (括号内)

σ_b	n	$\sigma_\varepsilon = 0.5$	$\sigma_\varepsilon = 1$	$\sigma_\varepsilon = 1.5$
		Mean(SD)	Mean(SD)	Mean(SD)
0.6	30	0.99707(0.00173)	0.99193(0.00381)	0.98194(0.00910)
	50	0.99848(0.00102)	0.99498(0.00268)	0.98942(0.00587)
	100	0.99929(0.00039)	0.99778(0.00110)	0.99479(0.00279)
	150	0.99951(0.00027)	0.99830(0.00136)	0.99620(0.00289)
1.2	30	0.99455(0.00500)	0.98892(0.00804)	0.98069(0.01078)
	50	0.99724(0.00314)	0.99400(0.00320)	0.98793(0.00638)
	100	0.99865(0.00143)	0.99683(0.00204)	0.99442(0.00275)
	150	0.99906(0.00128)	0.99743(0.00290)	0.99558(0.00341)
2	30	0.98986(0.01153)	0.98106(0.01753)	0.97233(0.01788)
	50	0.99350(0.00898)	0.98979(0.00909)	0.98403(0.01186)
	100	0.99679(0.00489)	0.99495(0.00548)	0.99217(0.00598)
	150	0.99763(0.00346)	0.99606(0.00404)	0.99470(0.00463)

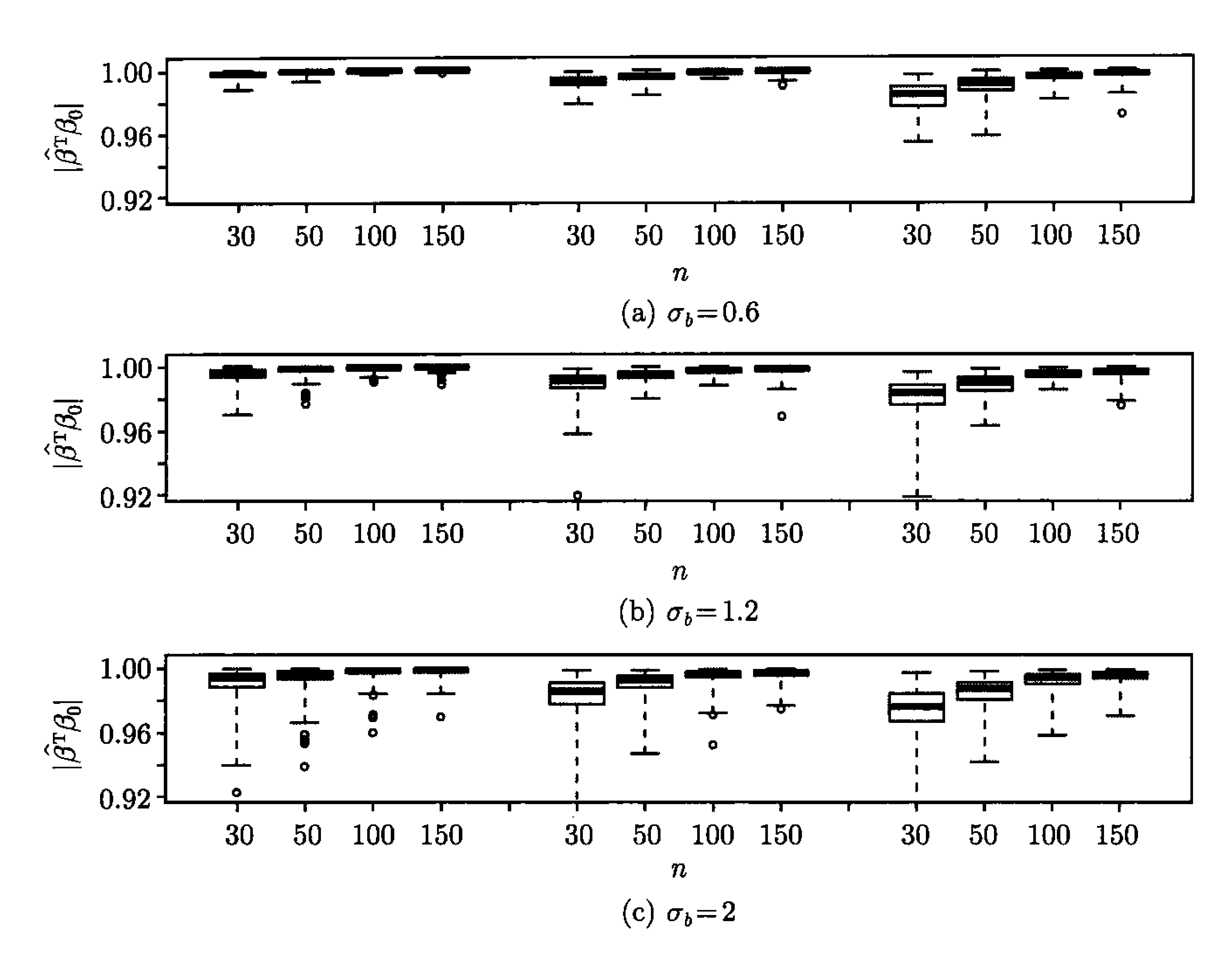


图 9.4.3 例 9.4.2 的模拟结果. 对每个 σ_b、样本量 $n = 30, 50, 100, 150$ 和 $\sigma_\varepsilon = 0.5, 1, 1.5$ 的绝对内积 $|\hat{\beta}^{\mathrm{T}}\beta_0|$ 的箱线图. (a) 对 $\sigma_b = 0.6$, (b) 对 $\sigma_b = 1.2$, (c) 对 $\sigma_b = 2$

例 9.4.3 为了进一步说明我们的算法在高维情形的适应性, 考虑单指标混合效应模型

$$Y_{ij} = 8\sin(\pi\beta_0^{\mathrm{T}}X_{ij}/2) + b_i + \varepsilon_{ij},$$

其中 $\beta_0 = (1/\sqrt{5}, 2/\sqrt{5}, 0, \cdots, 0)^{\mathrm{T}}$, $p = 30, 50, 80$, X_{ij} 由 $N(2, I_p)$ 分布产生, $g(u) = 8\sin(\pi u/2)$, b_i 和 ε_{ij} 都服从均值为 0 且方差分别为 1.44 和 1 的正态分布. 个体数目 n 分别取为 50, 100 和 150, 而每个个体重复观测次数 m 取为 3.

取权函数 $w(u) = I_{[-6,6]}(u)$, 用绝对内积 $|\hat{\beta}^{\mathrm{T}}\beta_0|$ 来度量估计量 $\hat{\beta}$ 的精度. 表 9.4.3 给出的结果是基于 200 次的重复运行. 在这个模拟中, 一个重要现象是所提出的方法在参数维数等于或略大于个体数目时仍能奏效. 另外, 内积的精度随 p 的增加而降低.

表 9.4.3 例 9.4.3 的模拟结果. 估计值 $\hat{\beta}$ 和真实值 β_0 的内积的均值和标准差 (括号内)

n	$p = 30$	$p = 50$	$p = 80$
	Mean(SD)	Mean(SD)	Mean(SD)
50	0.99573(0.00186)	0.98895(0.00450)	0.95493(0.02158)
100	0.99837(0.00058)	0.99681(0.00119)	0.99289(0.00230)
150	0.99894(0.00039)	0.99822(0.00056)	0.99663(0.00088)

9.4.2 一个实例

我们现在使用来自癫痫病研究 (Thall and Vail, 1990; Wang et al., 2005; Bai et al., 2009) 的实际数据集来说明本章提出的方法. 在实验阶段, 允许两个不同的方案 (安慰剂和抗痫药物) 处理 59 个癫痫患者. 患者被随机的安排接受其中的一种方案. 其中 28 个患者被使用安慰剂, 31 名患者使用抗痫药物治疗. 患者每隔两周到诊所做 4 次连续的检查, 报告过去两周癫痫发作的次数. 在这项研究中, 响应变量是每两周中癫痫的次数, 协变量分别为以年计的年龄的对数, 和基准癫痫数量 (先除以 4 再作对数变换). 一个科学的问题是药物能否有助于减少癫痫的发病次数. 因此, 我们使用本章的模型和方法来分析这个癫痫病数据, 比较两组患者的发病情况.

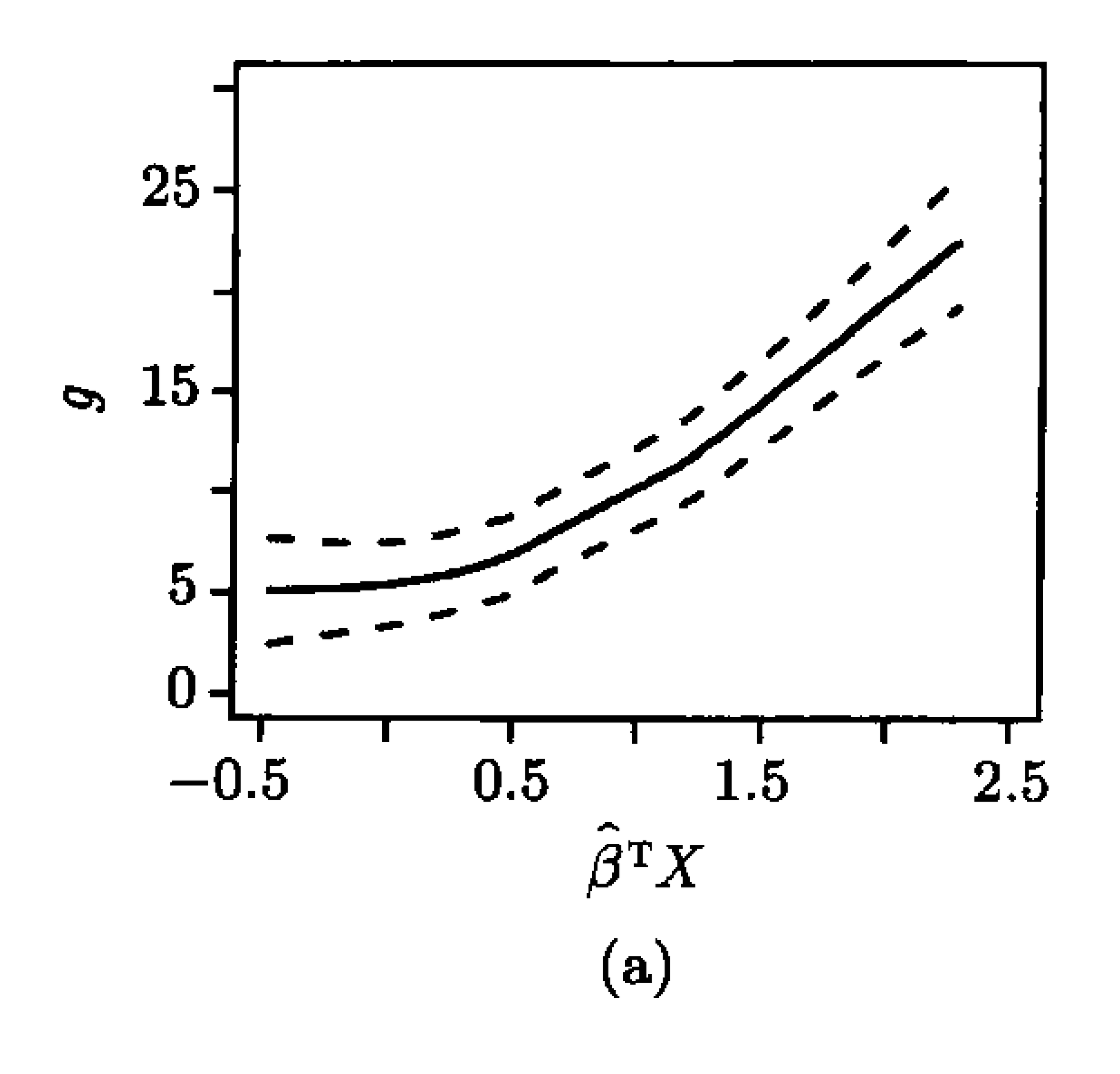
我们使用线性模型选择 β_0 的初始值, 利用 (9.2.5)~(9.2.9) 式来计算 β_0, σ_b^2 和 σ_ε^2 的估计值. 计算结果在表 9.4.4 中给出.

表 9.4.4 癫痫病数据的应用. β_1 和 β_2 以及 σ_b^2 和 σ_ε^2 的估计和标准误 (括号内)

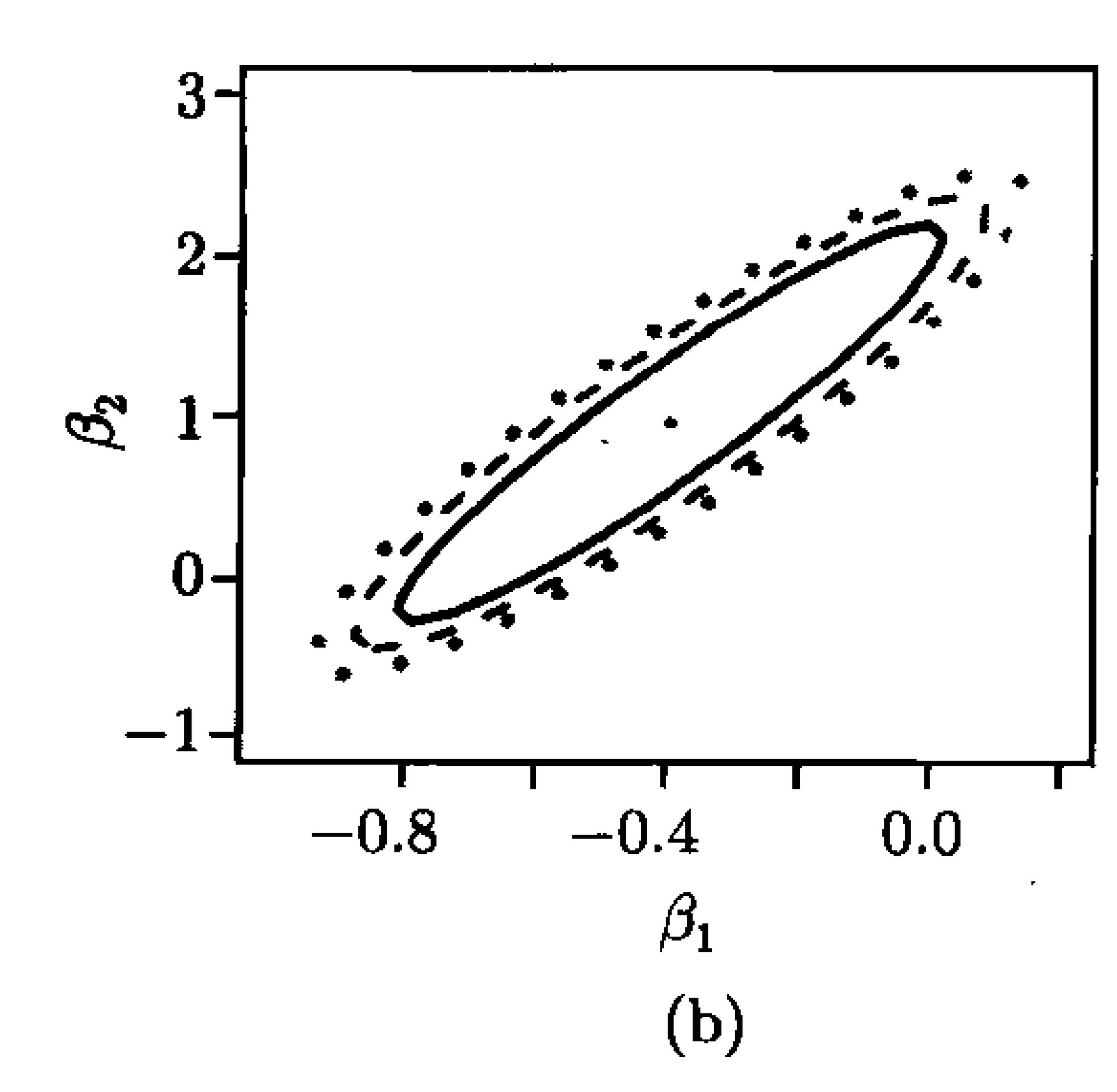
组	$\hat{\beta}_1$(SE)	$\hat{\beta}_2$(SE)	$\hat{\sigma}_b^2$	$\hat{\sigma}_\varepsilon^2$
安慰剂	−0.2651(0.5797)	0.9642(0.1594)	2.6988	72.0863
药物治疗	−0.0208(0.4845)	0.9998(0.0101)	0	64.0397
全体患者	−0.2670(0.1922)	0.9637(0.0533)	0	70.2278

由表 9.4.4 可以发现, 与使用安慰剂组相比, 抗痫药物组具有较大指标系数和较小的误差方差, 且由抗痫药物组得到的模型没有随机效应. 这个结果也表明了年龄和基准癫痫次数具有显著的统计意义. 除此以外, 由整个数据集获得的指标系数接近由安慰剂组获得的指标系数. 我们也计算了 β_0 的置信域, $g(\cdot)$ 的估计曲线和逐点的置信区间, 其结果如图 9.4.4 所示.

图 9.4.4(a),(c) 和 (e) 表明癫痫次数的预测均值随 $\hat{\beta}^{\mathrm{T}}X$ 的增加而增加. 在 (c) 中, 曲线在点 2.8837 处迅速增加. 正如 Thall 和 Vail(1990) 指出的那样, 药物组患者 #207 在基准线上和治疗后有极高的癫痫次数, 如图 9.4.5(a) 所示. 删除该患者的数据会得到预测值的明显下降, 尽管组内相关性仍比较大, 如图 9.4.5(b) 所示. 由



(a)



(b)

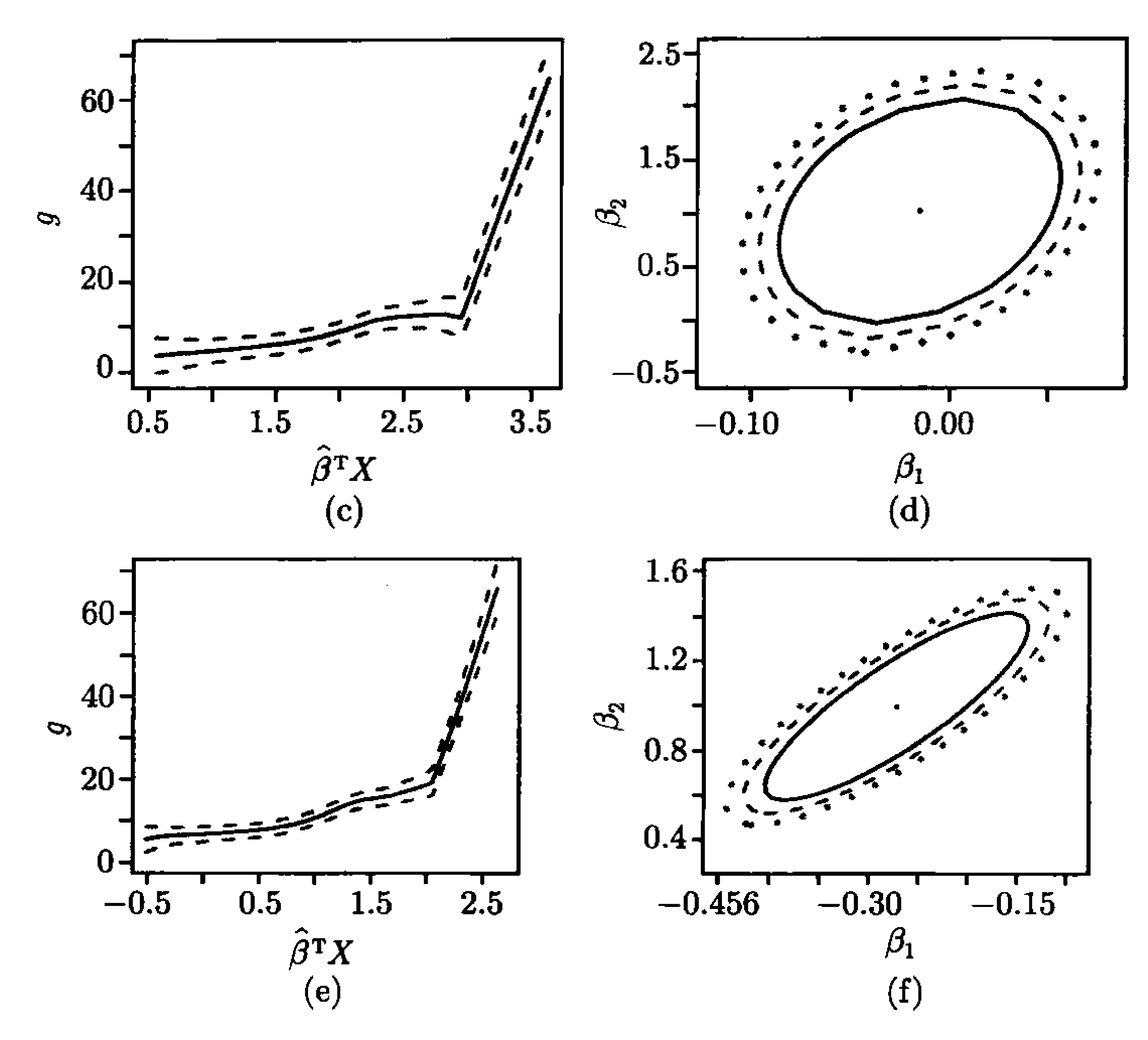


图 9.4.4 癫痫病数据的应用. 估计的曲线 (实线) 和 $g(\cdot)$ 的近似 0.95 逐点置信区间 (虚线); 以及 β_0 的近似 0.9(实线), 0.95(虚线) 和 0.99(点线) 置信域. (a) 和 (b) 对安慰剂组; (c) 和 (d) 对抗痫药物组; (e) 和 (f) 对所有数据

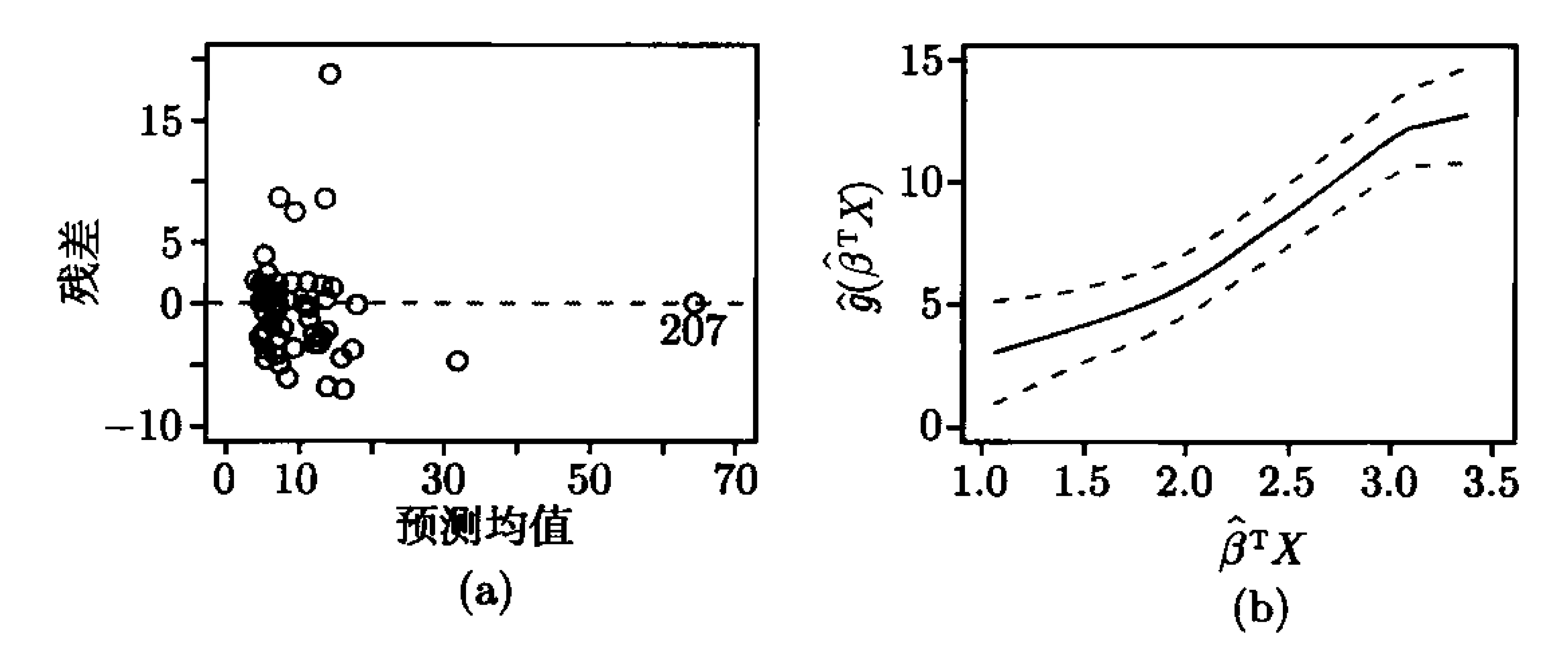


图 9.4.5 癫痫病数据的应用. (a) 癫痫病数的对数的残差 $\bar{Y}_i - \bar{\hat{g}}_i$, 预测均值为 $\bar{\hat{g}}_i$. (b) 对安慰剂组: 估计的曲线 (实线) 和 $g(\cdot)$ 的近似 0.95 逐点置信区间 (虚线)

拟氨基丁酸药物组和安慰组的调整的估计量的比较, 表明了治疗效果非常明显, 即说明药物可以帮助减少癫痫的发病次数. 然而, 对该患者数据的删除没有医学意义. 除此以外, 图 9.4.4(e) 和 (f) 表明了由整个数据估计的联系函数接近于抗痫药物组的估计曲线, 尽管指标系数接近于安慰剂组的指标系数. 这又表明了这两个组的治疗效果是有区别的.

9.5 定理的证明

本节中使用 c 表示任何正的常数, 每次出现时可以取不同值. 下列引理 9.5.1 可以直接由 Mack 和 Silverman(1982) 的结果推出.

引理 9.5.1 设 $(\xi_1,\eta_1),\cdots,(\xi_n,\eta_n)$ 是 iid 随机向量, 其中 η_i 是一维随机变量. 进一步, 假设 $E|\eta_1|^s<\infty$ 和 $\sup\limits_x\int|y|^sf(x,y)\mathrm{d}y<\infty$, 其中 $f(\cdot,\cdot)$ 表示 (ξ_1,η_1) 的联合密度函数. 设 $K(\cdot)$ 是具有有界支撑的有界正函数, 并满足 Lipschitz 条件. 如果对某个 $\epsilon<1-s^{-1}$, 有 $n^{2\epsilon-1}h\to\infty$, 则

$$\sup_x\left|\frac{1}{n}\sum_{i=1}^{n}\{K_h(\xi_i-x)\eta_i-E[K_h(\xi_i-x)\eta_i]\}\right|=O_P\left(\left[\frac{\log(1/h)}{nh}\right]^{1/2}\right),$$

定理 9.3.1 的证 利用引理 9.5.1, 并类似于定理 7.2.1 的证明, 可以证得: 对 $u\in\mathcal{U}_w$ 和 $\beta\in\mathcal{B}_n$ 一致成立

$$\begin{pmatrix}\hat{g}(u;\beta)-g(u)\\ h[\hat{g}'(u;\beta)-g'(u)]\end{pmatrix}=S^{-1}(u)\xi_n^*(u;\beta)+\frac{1}{2}h^2\begin{pmatrix}\mu_2g''(u)\\ \dfrac{\mu_3}{\mu_2}g''(u)\end{pmatrix}+o_P(c_n),\quad(9.5.1)$$

其中 $c_n=n^{-1/2}+h^2$, $S(u)=\sum\limits_{j=1}^{m}f_j(u)\otimes\mathrm{diag}(1,\mu_2)$, $\otimes$ 为 Kronecker 乘积, $\xi_n^*(u;\beta)=\left(\xi_{n,0}^*(u;\beta),\xi_{n,1}^*(u;\beta)\right)^{\mathrm{T}}$,

$$\xi_{n,l}^*(u;\beta)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\{Y_{ij}-g^{\mathrm{T}}(\beta^{\mathrm{T}}X_{ij})\}\left(\frac{\beta^{\mathrm{T}}X_{ij}-u}{h}\right)^jK_h(\beta^{\mathrm{T}}X_{ij}-u).$$

由引理 9.5.1 可以证得: 对 $u\in\mathcal{U}_w$ 和 $\beta\in\mathcal{B}_n$ 一致成立

$$\xi_n^*(u;\beta)=O_P\left(\left[\frac{\log(1/h)}{nh}\right]^{1/2}+n^{-1/2}\right).\qquad(9.5.2)$$

结合 (9.5.1) 和 (9.5.2) 式即可证得定理 9.3.1.

定理 9.3.2 的证 由 (9.5.1) 式可得

$$\hat{g}(u;\beta)-g(u)=\left[\sum_{j=1}^{m}f_j(u)\right]^{-1}\xi_{n,0}^*(u;\beta)+\frac{1}{2}h^2\mu_2g''(u)+o_P(c_n).\qquad(9.5.3)$$

注意到 $||\beta-\beta_0||=O(n^{-1/2})$, 可以证明

$$\xi_{n,0}^*(u;\beta)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}(Z_{ij}^{\mathrm{T}}b_i+\varepsilon_{ij})K_h(\beta_0^{\mathrm{T}}X_{ij}-u)+O_P(n^{-1/2}).$$

由 (Masry and Tjøstheim, 1995) 的定理 4.4, 可得

$$\sqrt{nh}\xi_{n,0}^*(u;\beta) \xrightarrow{D} N\left(0,\gamma^2(u)\right). \tag{9.5.4}$$

因此, 由 (9.5.3) 和 (9.5.4) 式即可完成定理 9.3.2 的证明.

定理 9.3.3 的证　该定理的证明分为五部分. 步骤 (i) 给出了整个证明的概述. 步骤 (ii)~(v) 给出了在步骤 (i) 中所用到的各种残差估计的证明.

步骤(i) 可以证明: 依概率存在满足 $\hat\beta^{(r)}-\beta_0^{(r)}=O_P(n^{-1/2})$ 的方程 $Q_n(\hat G,\beta^{(r)})=0$ 的解 $\hat\beta^{(r)}$. 论证的细节类似于 Wang 等 (2010) 中定理 2, 此处省略其证明. 对某个正的常数 $c_2>0$, 设 $\mathcal{B}_n^{(r)}=\{\beta^{(r)}:||\beta^{(r)}-\beta_0^{(r)}||\leqslant c_2n^{-1/2}\}$. 记 $\mathcal{G}=\{\widetilde G:\mathcal{U}_w^m\times\mathcal{B}\mapsto\mathbf{R}^m\}$ 和 $||\widetilde G||_{\mathcal{G}}=\sup_{u\in\mathcal{U}_w^m,\beta^{(r)}\in\mathcal{B}_n^{(r)}}||\widetilde G(u;\beta)||$, 其中 $\beta=\beta(\beta^{(r)})$. 因为当 $||\beta||=1$ 时, $||\beta-\beta_0||\leqslant c_1n^{-1/2}$ 与 $||\beta^{(r)}-\beta_0^{(r)}||\leqslant c_2n^{-1/2}$ 是等价的, 由定理 9.3.1 知 $||\hat G-G||_{\mathcal{G}}=o_P(1)$ 和 $\max|\hat G'_\Delta-G'_\Delta|_{\mathcal{G}}=o_P(1)$, 其中 $\max|H|$ 表示矩阵 H 中每个元素的绝对值的最大值. 因此, 对 $\delta=\delta_n\to0$ 和 $\delta>0$, 可以假设 $\mathcal{G}_\delta\in\hat G$, 其中

$$\mathcal{G}_\delta=\{\widetilde G\in\mathcal{G}:||\widetilde G-G||_{\mathcal{G}}\leqslant\delta,\max|\widetilde G'_\Delta-G'_\Delta|_{\mathcal{G}}\leqslant\delta\}. \tag{9.5.5}$$

设

$$Q(G,\beta^{(r)})=\frac{1}{n}\sum_{i=1}^n E\left\{J_{\beta^{(r)}}^{\mathrm T}X_1^{\mathrm T}G'_\Delta(X_1\beta)W(X_1\beta)V_i^{-1}[Y_1-G(X_1\beta;\beta)]\right\},$$

其中 $G(X_1\beta;\beta)=(g(\beta^{\mathrm T}X_{11};\beta),\cdots,g(\beta^{\mathrm T}X_{1m};\beta))^{\mathrm T}$, $g(\beta^{\mathrm T}X_{1j};\beta)=E(Y_{1j}|\beta^{\mathrm T}X_{1j})$. 显然, $g(\beta_0^{\mathrm T}X_{1j};\beta_0)=g(\beta_0^{\mathrm T}X_{1j})$, $G(X_1\beta_0;\beta_0)=G(X_1\beta_0)$. 定义函数 $Q(\widetilde G,\beta^{(r)})$ 关于 $\widetilde G(\cdot;\beta)$ 在 $G(\cdot;\beta)$ 处沿方向 $\widetilde G(\cdot;\beta)-G(\cdot;\beta)$ 的导数 $\varpi(G(\cdot;\beta),\beta^{(r)})$:

$$\begin{aligned}&\varpi(G(\cdot;\beta),\beta^{(r)})(\widetilde G(\cdot;\beta)-G(\cdot;\beta))\\=&\lim_{\tau\to0}[Q(G(\cdot;\beta)+\tau(\widetilde G(\cdot;\beta)-G(\cdot;\beta)),\beta^{(r)})-Q(G(\cdot;\beta),\beta^{(r)})]\cdot\frac{1}{\tau}.\end{aligned}$$

那么

$$\begin{aligned}&\varpi(G(X_1\beta;\beta),\beta^{(r)})(\widetilde G(X_1\beta;\beta)-G(X_1\beta;\beta))\\=&-\frac{1}{n}\sum_{i=1}^n\int\mu(u)f^{\mathrm T}(u)G'_\Delta(u)W(u)V_i^{-1}[\widetilde G(u;\beta)-G(u;\beta)]\mathrm{d}u,\end{aligned} \tag{9.5.6}$$

其中 $\mu(u)=E\{J_{\beta^{(r)}}^{\mathrm T}X_1|X_1\beta=u\}$ 和 $f(u)=(f_1(u_1),\cdots,f_m(u_m))^{\mathrm T}$.

简单的计算, 可以将 $Q_n(\hat G,\beta^{(r)})$ 分解为

$$Q_n(\hat G,\beta^{(r)})=M_1(\beta^{(r)})+M_2(\beta^{(r)})+M_3(\beta^{(r)})+M_4(\beta_0^{(r)})+Q(G,\beta^{(r)}), \tag{9.5.7}$$

其中 $Q_n(\hat{G},\beta^{(r)})$ 在 (9.2.2) 中定义,

$$\begin{aligned} M_1(\beta^{(r)}) =& Q_n(\hat{G},\beta^{(r)}) - Q(\hat{G},\beta^{(r)}) - Q_n(G,\beta_0^{(r)}), \\ M_2(\beta^{(r)}) =& Q(\hat{G},\beta^{(r)}) - Q(G,\beta^{(r)}) \\ &-\varpi(G(X_1\beta;\beta),\beta^{(r)})[\hat{G}(X_1\beta;\beta) - G(X_1\beta;\beta)], \\ M_3(\beta^{(r)}) =& \varpi(G(X_1\beta;\beta),\beta^{(r)})[\hat{G}(X_1\beta;\beta) - G(X_1\beta;\beta)] \\ &-\varpi(G(X_1\beta_0),\beta_0^{(r)})[\hat{G}(X_1\beta_0;\beta_0) - G(X_1\beta_0)], \end{aligned}$$

$$M_4(\beta_0^{(r)}) = Q_n(G,\beta_0^{(r)}) + \varpi(G(X_1\beta_0),\beta_0^{(r)})[\hat{G}(X_1\beta_0;\beta_0) - G(X_1\beta_0)].$$

步骤 (ii)~(v) 证明了下列结果:

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^{(r)}} ||M_1(\beta^{(r)})|| = o_P(n^{-1/2}), \tag{9.5.8}$$

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^{(r)}} ||M_2(\beta^{(r)})|| = o_P(n^{-1/2}), \tag{9.5.9}$$

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^{(r)}} ||M_3(\beta^{(r)})|| = o(n^{-1/2}), \tag{9.5.10}$$

$$\sqrt{n}M_4(\beta_0^{(r)}) \xrightarrow{D} N(0,A), \tag{9.5.11}$$

其中 A 在定理 9.3.3 中给出定义.

记 $Q(G,\beta_0^{(r)})=0$, 由 Taylor 公式, 对 $\beta^{(r)}\in\mathcal{B}_n^{(r)}$ 一致成立

$$Q(G,\beta^{(r)}) = -A(\beta^{(r)}-\beta_0^{(r)}) + o_P(n^{-1/2}). \tag{9.5.12}$$

由 (9.5.7)~(9.5.10) 和 (9.5.12) 式可得, 对 $\beta^{(r)}\in\mathcal{B}_n$ 一致成立

$$Q_n(\hat{G},\beta^{(r)}) = M_4(\beta_0^{(r)}) - A(\beta^{(r)}-\beta_0^{(r)}) + o_P(n^{-1/2}).$$

注意到 $Q_n(\hat{G},\hat{\beta}^{(r)})=0$, 由上述方程可得

$$\hat{\beta}^{(r)} - \beta_0^{(r)} = A^{-}M_4(\beta_0^{(r)}) + o_P(n^{-1/2}). \tag{9.5.13}$$

容易证明

$$\hat{\beta} - \beta_0 = J_{\beta_0^{(r)}}(\hat{\beta}^{(r)} - \beta_0^{(r)}) + O_P(n^{-1}). \tag{9.5.14}$$

由 (9.5.13) 和 (9.5.14) 式可得

$$\sqrt{n}(\hat{\beta}-\beta_0) = \sqrt{n}\,J_{\beta_0^{(r)}}A^{-}M_4(\beta_0^{(r)}) + o_P(1).$$

结合 (9.5.11) 式即可完成定理 9.3.3 的证明.

步骤(ii) 我们返回来证明 (9.5.8) 式. 对任意$\widetilde{G}$, 记$r_n(\widetilde{G},\beta^{(r)})=\sqrt{n}[Q_n(\widetilde{G},\beta^{(r)})-Q(\widetilde{G},\beta^{(r)})]$. 注意到 $Q(G,\beta_0^{(r)})=0$, 显然有

$$M_1(\beta^{(r)})=n^{-1/2}[r_n(\hat{G},\beta^{(r)})-r_n(G,\beta_0^{(r)})]. \tag{9.5.15}$$

因此, 可以证明经验过程 $\{r_n(\widetilde{G},\beta^{(r)}):\widetilde{G}\in\mathcal{G}_1,\beta^{(r)}\in\mathcal{B}_1^{(r)}\}$ 具有随机等度连续性, 其中 $\mathcal{B}_1^{(r)}=\{\beta^{(r)}:||\beta^{(r)}-\beta_0^{(r)}||\leqslant c_2\}$, $\mathcal{G}_1$ 在 (9.5.5) 式中定义, $\delta=1$, 他们分别是 $\mathcal{B}$ 和 $\mathcal{G}$ 的子集. 由 $\delta\to 0$ 知当 n 足够大时 $\delta<1$, 这足可用来证明 (9.5.8) 式. 通过验证 (Doukhan et al., 1995) 中的定理 1 的条件可得到上述随机等度连续性. 因此, 对 $\widetilde{G}\in\mathcal{G}_1$ 和 $\beta^{(r)}\in\mathcal{B}_1$ 一致成立 $r_n(\widetilde{G},\beta^{(r)})-r_n(G,\beta_0^{(r)})=o_P(1)$. 结合 (9.5.15) 式即可完成 (9.5.8) 式的证明.

步骤 (iii) 现在证明 (9.5.9) 式. 由 (9.5.6) 式可得

$$\begin{aligned}M_2(\beta^{(r)})=&-\int\mu(u)f^{\mathrm{T}}(u)\{\hat{G}'_\Delta(u;\beta)-G'_\Delta(u)\}W(u)V^{-1}\\&\times[\hat{G}(u;\beta)-G(u)]\mathrm{d}u+o_P(n^{-1/2}).\end{aligned} \tag{9.5.16}$$

利用定理 9.3.1 和条件 (C9.3.5) 容易完成 (9.5.9) 式的证明.

步骤 (iv) 现在证明 (9.5.10) 式. 记 $\varphi(\beta^{(r)})=\hat{G}(X_1\beta;\beta)-G(X_1\beta;\beta)$, $\psi(\beta^{(r)})=J_{\beta^{(r)}}^{\mathrm{T}}X_1^{\mathrm{T}}G'_\Delta(X_1\beta)W(X_1\beta)V_i^{-1}$. 由 (9.5.6) 式可以推出

$$\begin{aligned}M_3(\beta^{(r)})&=-E[\psi(\beta^{(r)})\varphi(\beta^{(r)})]+E[\psi(\beta_0^{(r)})\varphi(\beta_0^{(r)})]\\&=-E\{\psi(\beta^{(r)})[\varphi(\beta^{(r)})-\varphi(\beta_0^{(r)})]\}\\&\quad-E\{[\psi(\beta^{(r)})-\psi(\beta_0^{(r)})]\varphi(\beta_0^{(r)})\}\\&\equiv M_{31}(\beta^{(r)})+M_{32}(\beta^{(r)}).\end{aligned}$$

可以证明

$$\sup_{\beta^{(r)}\in\mathcal{B}_n^{(r)}}||M_{3\nu}(\beta^{(r)})||=o(n^{-1/2}),\quad \nu=1,2.$$

因此, (9.5.10) 式得证.

步骤 (v) 最后证明 (9.5.11) 式. 由 (9.5.1) 式可知, 对 $u\in\mathcal{U}_w^m$ 和 $\beta\in\mathcal{B}_n^{(r)}$ 一致成立

$$\hat{G}(u;\beta)-G(u)=[f^{\mathrm{T}}(u)1_m]^{-1}\tilde{\xi}_n^*(u;\beta)+\frac{1}{2}h^2\mu_2G''(u)+o_P(c_n), \tag{9.5.17}$$

其中 $c_n=n^{-1/2}+h^2$, $\tilde{\xi}_n^*(u;\beta)=(\xi_{n,0}^*(u_1;\beta),\cdots,\xi_{n,0}^*(u_m;\beta))^{\mathrm{T}}$, $G''(u)=(g''(u_1),\cdots,g''(u_m))^{\mathrm{T}}$. 记 $C(u)=E[J_{\beta_0^{(r)}}^{\mathrm{T}}X_1^{\mathrm{T}}G'_\Delta(X_1\beta_0)W(X_1\beta_0)V^{-1}|X_1\beta_0=u]$, 用 $f_0(u)=$

$(f_{01}(u_1),\cdots,f_{0m}(u_m))^{\mathrm{T}}$ 表示 $X_1\beta_0$ 的联合密度函数. 由 (9.5.6) 和 (9.5.17) 式以及控制收敛定理 (Loève, 2000), 可以证得

$$\begin{aligned}&\varpi(G(X_1\beta_0),\beta_0^{(r)})[\hat{G}(X_1\beta_0;\beta_0)-G(X_1\beta_0)]\\=&-\int C(u)[\hat{G}(u;\beta_0)-G(u)]f_0(u)\mathrm{d}u\\=&-\frac{1}{n}\sum_{i=1}^{n}C(X_i\beta_0)(Z_ib_i+\varepsilon_i)+o_P(n^{-1/2}).\end{aligned}$$

结合 (9.2.2) 式和 $M_4(\beta_0^{(r)})$ 的定义即可证得

$$M_4(\beta_0^{(r)})=\frac{1}{n}\sum_{i=1}^{n}\zeta_i(Z_ib_i+\varepsilon_i)+o_P(n^{-1/2}),$$

其中 $\zeta_i=J_{\beta_0^{(r)}}^{\mathrm{T}}X_i^{\mathrm{T}}G'_{\Delta}(X_i\beta_0)W(X_i\beta_0)V_i^{-1}-C(X_i\beta_0)$. 因此, 由独立样本的中心极限定理和 Slutsky 定理, 即可以得到 (9.5.11) 式.

定理 9.3.4 的证 首先证明 (9.3.1). 记 $\bar{\varepsilon}_i=m^{-1}\sum_{j=1}^{m}\varepsilon_{ij}$, $\tilde{g}_{ij}=g(X_{ij}^{\mathrm{T}}\beta_0)-\hat{g}(X_{ij}^{\mathrm{T}}\hat{\beta};\hat{\beta})$, $\bar{\tilde{g}}=m^{-1}\sum_{j=1}^{m}\tilde{g}_{ij}$. 由 (9.2.7) 式可得

$$\begin{aligned}\hat{\sigma}_{\varepsilon}^2=&\frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}[(\varepsilon_{ij}-\bar{\varepsilon}_i)+(\tilde{g}_{ij}-\bar{\tilde{g}}_i)]^2\\=&\frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(\varepsilon_{ij}-\bar{\varepsilon}_i)^2+\frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(\tilde{g}_{ij}-\bar{\tilde{g}}_i)^2\\&+\frac{2}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(\varepsilon_{ij}-\bar{\varepsilon}_i)(\tilde{g}_{ij}-\bar{\tilde{g}}_i)\\\equiv&L_1+L_2+L_3.\end{aligned}$$

由大数定律以及定理 9.3.1 和定理 9.3.3, 可以证明 $L_1=\sigma_{\varepsilon}^2+O_P(n^{-1/2})$, $L_2=O_P(n^{-1/2})$, $L_3=O_P(n^{-1/2})$, 因此, (9.3.1) 式成立. 对估计量 (9.2.9), 类似地也可以证明 (9.3.1) 式成立.

现在证明 (9.3.2) 式. 简单计算可得

$$\begin{aligned}\hat{\sigma}_b^2=&\frac{1}{n}\sum_{i=1}^{n}(b_i+\bar{\varepsilon}_i+\bar{\tilde{g}}_i)^2-\hat{\sigma}_{\varepsilon}^2/m\\=&\frac{1}{n}\sum_{i=1}^{n}(b_i+\bar{\varepsilon}_i)^2+\frac{1}{n}\sum_{i=1}^{n}\bar{\tilde{g}}_i^2+\frac{2}{n}\sum_{i=1}^{n}(b_i+\bar{\varepsilon}_i)\bar{\tilde{g}}_i-\hat{\sigma}_{\varepsilon}^2/m\\\equiv&J_1+J_2+J_3-\hat{\sigma}_{\varepsilon}^2/m.\end{aligned}$$

不难证明 $J_1 = \sigma_b^2 + \sigma_\varepsilon^2/m + O_P(n^{-1/2})$, $J_2 = O_P(n^{-1/2})$ 和 $J_3 = O_P(n^{-1/2})$. 结合定理 9.3.4 的第一个公式即可完成 (9.3.2) 式的证明.

参考文献

Arnold S F. 1981. The Theory of Linear Models and Multivariate Analysis. New York: John Wiley & Sons

Bai Y, Fung W K, Zhu Z Y. 2009. Penalized quadratic inference functions for single-index models with longitudinal data. Journal of Multivariate Analysis, 100: 152～161

Cai T, Cheng S C, Wei L J. 2002. Semiparametric mixed-effects models for clustered failure time data. Journal of the American Statistical Association, 97: 514～522

Carro J M. 2007. Estimating dynamic panel data discrete choice models with fixed effects. Journal of Econometrics, 140: 503～528

Carroll R J, Fan J, Gijbels I, et al.. 1997. Generalized partially linear single-index models. Journal of the American Statistical Association, 92: 477～489

Cui X, Härdle W, Zhu L X. 2011. The EFM approach for single-index models. The Annals of Statistics, 39(3): 1658～1688

Doukhan P, Massart P, Rio E. 1995. Invariance principles for absolutely regular empirical processes. Annals of Institute H Poincaré Probability Statistics 31: 393～427

Fan J, Gijbels I. 1996. Local Polynomial Modeling and Its Applications. London: Chapman and Hall

Gu C, Ma P. 2005. Optimal smoothing in nonparametric mixed-effect models. The Annals of Statistics, 33: 1357～1379

Hall P, Maiti T. 2006. Nonparametric estimation of mean-squared prediction error in nested-error regression models. The Annals of Statistics, 34: 1733～1750

Härdle W, Hall P, Ichimura H. 1993. Optimal smoothing in single-index models. The Annals of Statistics, 21: 157～178

Harville D A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. with discussion. Journal of the American Statistical Association, 72: 320～340

Honorá B, Kyriazidou E. 2000. Panel data discrete choice models with lagged dependent variables. Econometrica, 68: 839～874

Hristache M, Juditsky A, Spokoiny V. 2001. Direct estimation of the index coefficient in a single-index model. The Annals of Statistics, 29: 595～623

Ichimura H. 1993. Semiparametric least squares. SLS. and weighted SLS estimation of single-index models. Journal of Econometrics, 58: 71～120

Jiang J. 2007. Linear and Generalized Linear Mixed Models and Their Applications. New

York: Springer

Ke C L, Wang Y D. 2001. Semiparametric nonlinear mixed-effects models and their applications(with discussion). Journal of the American Statistical Association, 96: 1272~1298

Li K C. 1991. Sliced inverse regression for dimension reduction. with discussion. Journal of the American Statistical Association, 86: 316~342

Liang K Y, Zeger S L. 1986. Longitudinal data analysis using generalized linear models. Biometrika, 73: 13~22

Lin X H, Carroll R J. 2000. Nonparametric function estimation for clustered data when the predictor is measured without/with error. Journal of the American Statistical Association, 95: 520~534

Lindstrom M J, Bates D M. 1990. Nonlinear mixed-effects models for repeated measures data. Biometrics, 46: 673~687

Loève M. 2000. Probability Theory I, 4th Edition. New York: Springer-Verlag

Mack Y P, Silverman B W. 1982. Weak and strong uniform consistency of kernel regression estimates, Z Wahrscheinlichkeitstheorie verw Gebiete, 61: 405~415

Masry E, Tjøstheim D. 1995. Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. Econometric Theory, 11: 258~289

Naik P, Tsai C L. 2000. Partial least squares estimator for single-index. Journal of the Royal Statistical Society, Series B, 62: 763~771

Pang Z, Xue L G. 2012. Estimation for the single-index models with random effects. Computational Statistics and Data Analysis, 56: 1837~1853

Ruckstuhl A F, Welsh A H, Carroll R J. 2000. Nonparametric function estimation of the relationship between two repeatedly measured variables. Statistica Sinica, 10: 51~71

Thall P, Vail S C. 1990. Some covariance models for longitudinal count data with over dispersion. Biometrics, 46: 657~671

Wang J L, Xue L G, Zhu L X, et al.. 2010. Partial-linear single-index model with noised variable. The Annals of Statistics, 38: 246~272

Wang Y G, Lin X, Zhu M. 2005. Robust estimating functions and bias correction for longitudinal data analysis. Biometrics, 61: 684~691

Wu H L, Zhang J T. 2002. Local polynomial mixed-effects models for longitudinal data. Journal of the American Statistical Association, 97: 883~897

Xia Y C. 2006. Asymptotic distributions for two estimators of the single-index model. Econometric Theory, 22: 1112~1137

Xia Y C, Li W K. 1999. On single-index coefficient regression models. Journal of the American Statistical Association, 94: 1275~1285

Xia Y C, Tong H, Li W K, et al.. 2002. An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society, Series B, 64: 363~410

Xue L G. 2010. Empirical likelihood local polynomial regression analysis of clustered data. Journal of Scandinavain Statistics, 37: 644～663

Xue L G, Zhu L X. 2006. Empirical likelihood for single-index models. Journal of Multivariate Analysis, 97: 1295～1312

Zeger S L, Diggle P J. 1994. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics, 50: 689～699

第10章　可加模型

10.1　引　　言

非参数回归方法在统计学中发挥了重要作用, 由于它的灵活性而在统计分析中被广泛使用. 研究者在进行数据分析中不需要假定响应变量与协变量之间相关性的形式. 然而, 当协变量的个数增多时非参数回归方法会难以执行. 尤其多于两个协变量时估计的精度会降低. 可加模型提供了解决这个问题的一个有效方法. Friedman 和 Stuetzle(1981) 最早提出了可加模型, 后来 Hastie 和 Tibshirani(1990) 在他们的著作中对可加模型的理论和方法作了详细阐述. 假定响应变量 Y 关于协变量 X 的条件期望可以表示为

$$E[Y|X=(x_1,\cdots,x_d)]=m_1(x_1)+\cdots+m_d(x_d), \tag{10.1.1}$$

则称它为可加模型. 由于模型 (10.1.1) 将非参数回归中的多维光滑削减为一维光滑问题, 因此它能够分析难以处理的多维问题. 模型 (10.1.1) 的主要优点是它避免了多维数据中经常遇到的 "维数灾祸".

随着实际应用的需要, 对可加模型的理论研究也得到不断发展. Stone(1985) 证明了可加分量的参数估计量可达到一维非参数回归估计的最优收敛速度, 并且该速度不依赖于变量的个数. 后来, Stone(1994) 和 Huang(1998, 2003) 称 Stone(1985) 提出的估计为多项式样条估计. Xue 和 Yang(2006a) 进一步将这种估计方法推广到弱相依数据的情况. Buja 等 (1989) 描述了怎样的线性光滑可以使用后移算法来估计可加模型, 指出了后移算法正是解可加模型的正规方程组的 Gauss-Seidel 迭代方法. Tjøtheim 和 Auestad(1994) 与 Linton 和 Nielsen(1995) 首次提出了边际整合方法, 并证明了他们的方法对可加模型达到了与单变量局部线性估计同样的收敛速度 $O_P(n^{-2/5})$. 人们在后来的研究中发展了这种方法, 与之相关的文献可参见 (Linton and Hädle, 1996; Fan et al., 1998; Yang et al., 1999; Hädle 2003; 2006; Sperlich et al., 2002; Xue and Yang, 2006b) 等. Opsomer 和 Ruppert(1997) 利用局部多项式回归方法构造了二变量可加模型的估计量, 并在较强的假定下研究了所提出的估计量的渐近性质. Opsomer(2000) 对协变量个数大于 2 的一般情况推广了他们的理论结果. Mammen 等 (1999) 提出了一个具有很好理论性质的修正的后移算法, 后来被 Nielsen 和 Sperlich(2005) 所改良, 并称为光滑后移估计. Horowitz 和 Mammen(2004) 研究了具有未知联系函数的可加模型的估计问题. 当可加分量为二

次连续可微时, 他们所构造的估计量具有 $n^{-2/5}$ 的弱收敛速度渐近于正态分布, 无论协变量的维数如何该结果都成立, 这就避免了维数灾祸. Wang 和 Yang(2007) 研究了非线性可加自回归模型, 提出了分量函数的样条后移核估计量, 并建立了有关的理论结果. Wang 和 Yang(2009) 使用样条方法构造了可加模型中分量函数的一步后移估计量; 在较弱的条件下, 他们提出的估计量的逐点分布渐近等价于单变量核 (局部线性) 估计量, 因此在任何点上维数有效地降为 1, 并且在正态误差假定下降维在任何区间上一致成立. Lin(2009) 对可加模型提出了一个两阶段估计方法; 他们的方法能够呈现回归函数空间和收敛速度之间的相关性, 且证明了提出的估计量可以达到最优收敛速度. Ravikumar(2009) 对稀疏可加模型提出了一种新的估计方法. 他们获得了一种实际有效的拟合模型的算法, 即使协变量个数大于样本量时也是如此. 他们也给出了稀疏可加模型的理论性质和实际数据的分析. Alimadad 和 Salibian-Barrera(2011) 基于后移算法讨论了广义可加模型的 outlier- 稳健拟合问题, 其基本思想是使用从稳健拟似然方程获得的广义局部得分算法代替基于权的极大似然. 他们证明了所得到的估计的平均函数的统计性质. 其他有关的研究工作还包括 (Andrews and Whang, 1990; Kim et al., 1999; Hädle et al., 2001; Huang and Yang, 2004; Linton, 1997; You and Zhou, 2007; Debbarh and Maillot, 2008; Jiang and Li, 2008; Yu et al., 2008; Müller and Yao, 2008; Alvarez and Pardinas, 2009; Fahrmeir and Kneib, 2009; Xia, 2009; De Bock et al., 2010; Xue and Liang, 2010) 等.

一些作者研究了可加模型的模型选择和变量选择问题. Wood(2008) 对广义可加模型光滑度选择提出了一个计算上高效稳定的方法. Zheng(2008) 对高维可加模型提出了一种选择分量和光滑度的方法. 他们的方法的思想是对不同的分量采用不同的惩罚, 且在一个分量中的所有光滑参数采用同样的惩罚. Huang 等 (2010) 利用 B 样条提出可加模型的变量选择方法, 利用这样的逼近, 分量选择的问题变为选择一组展开系数的问题. 他们利用自适应组 LASSO 选择非零分量, 使用组 LASSO 得到初始估计且削减问题的维数, 并证明了相应的统计性质. 其他相关的成果还有 (Belitz and Lang, 2008; Panagiotelis and Smith, 2008; Cantoni et al., 2011; Marra and Wood, 2011) 等.

Opsomer 和 Ruppert(1999) 将可加模型推广到了可加部分线性模型. 他们提出了模型的参数分量 $\sqrt{n}$ 相合的后移估计量. 对单一非参数项的情况发展了估计量的渐近性质, 并推广到任意多个非参数可加项的情形. 他们通过极小化偏差和方差的渐近表达提出了一个最优带宽的估计量, 并提出了模型拟合与带宽选择的快速执行算法. Liang 等 (2009) 研究了可加部分线性模型的经验似然推断, 在线性协变量具有测量误差下构造了参数分量的经验对数似然比统计量, 并证明了该统计量渐近服从 χ^2 分布. Liang 等 (2008) 在线性协变量具有测量误差下考虑了可加部分线性模型的统计推断. 他们提出了兴趣参数的衰减校正和模拟外推 (SIMEX) 的估

计量, 在不需要欠光滑的情况下证明了提出的估计量的渐近正态性, 同时也讨论了 SIMEX 方法的渐近性质. Wang 等 (2011) 研究了广义可加部分线性模型, 提出对非参数函数的估计使用多项式样条光滑, 获得了线性参数的基于拟似然的估计量, 并证明了参数分量的估计量的渐近正态性. 他们提出的方法避免了求解像基于核方法的大系统方程的问题. 他们通过对线性参数利用非凸惩罚拟似然发展了一种变量选择方法, 并证明了它具有渐近 oracle 性质. Zhang 和 Liang(2011) 研究了广义可加部分线性模型中的模型选择和模型平均. 他们利用多项式样条逼近非参数函数, 证明了线性参数的估计量是渐近正态的. 他们发展了基于拟似然原理标准的聚焦信息准则 (FIC) 和频率模型平均 (FMA) 估计量, 并讨论了 FIC 和 FMA 的理论性质. 其他工作还有 (Fahrmeir and Lang, 2001; Manzan and Zerom, 2005; Wang et al., 2010; Biedermann et al., 2011; Liu et al., 2011; Zhou et al., 2011) 等.

10.2 估 计 方 法

设 $\{(X_i, Y_i), 1 \leqslant i \leqslant n\}$ 为来自 (X, Y) 的 $d+1$ 维 iid 随机样本. 考虑可加模型

$$Y_i = \theta + m_1(X_{i1}) + \cdots + m_d(X_{id}) + \varepsilon_i, \tag{10.2.1}$$

其中 θ 是截距项, X_{ij} 是 X_i 的第 j 个分量, 诸 m_j 是未知函数, 误差 $\{\varepsilon_i, 1 \leqslant i \leqslant n\}$ 是均值为 0 的 iid 随机变量. 为确保诸函数 m_j 的可识别性, 我们包含截距项 θ 且假设 $E[m_j(X_{ij})] = 0,\ j = 1, \cdots, d$.

本节主要阐述 θ 和诸 m_j 的估计方法. 本节主要内容的取材来自下面一些文献: 10.2.1 小节的内容取材于 (Opsomer and Ruppert, 1997; Opsomer, 2000); 10.2.2 小节的内容取材于 (Wang and Yang, 2007; Wang and Yang, 2009); 10.2.3 小节的内容取材于 (Stone, 1985; Xue and Yang, 2006a; Xue and Liang, 2010).

10.2.1 局部多项式估计

本节假定模型误差 ε_i 的方差为 σ^2. 首先引入一些记号. 记 $m(x_1, \cdots, x_d) = \theta + m_1(x_1) + \cdots + m_d(x_d)$, $X_i = (X_{i1}, \cdots, X_{id})^{\mathrm{T}}$, $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^{\mathrm{T}}$, $\boldsymbol{X}_j = (X_{1j}, \cdots, X_{nj})^{\mathrm{T}}$, $\boldsymbol{m}_j = (m_j(X_{1j}), \cdots, m_j(X_{nj}))^{\mathrm{T}}$, $j = 1, \cdots, d$. 对任何常数 c, 用 c 表示 n 维向量 $(c, \cdots, c)^{\mathrm{T}}$. 用 s_{j,x_j}^{T} 表示第 j 个协变量在 x_j 处局部多项式回归的等价核, 即

$$s_{j,x_j}^{\mathrm{T}} = e_j^{\mathrm{T}}(\breve{X}_j^{\mathrm{T}} W_j \breve{X}_j)^{-1} \breve{X}_j^{\mathrm{T}} W_j, \quad j = 1, \cdots, d,$$

其中 e_j 表示第 j 个位置的元素为 1 其余皆为 0 的向量,

$$W_j = \mathrm{diag}\left\{K_{h_j}(X_{1j} - x_j), \cdots, K_{h_j}(X_{nj} - x_j)\right\}, \quad j = 1, \cdots, d,$$

$K_{h_j}(\cdot) = h_j^{-1}K(\cdot/h_j)$, K 为某个核函数, h_j 为带宽,

$$\breve{X}_j = \begin{pmatrix} 1 & (X_{1j}-x_j) & \cdots & (X_{1j}-x_j)^{p_j} \\ \vdots & \vdots & & \vdots \\ 1 & (X_{nj}-x_j) & \cdots & (X_{nj}-x_j)^{p_j} \end{pmatrix}, \quad j=1,\cdots,d,$$

p_j 是拟合 m_j 的局部多项式的阶. 用 S_j 表示 m_j 在观测 $\boldsymbol{X}_j$ 处局部多项式拟合的光滑矩阵, 即

$$S_j = (s_{j,X_{1j}},\cdots,s_{j,X_{nj}})^{\mathrm{T}}, \quad j=1,\cdots,d.$$

在观测点处 $\boldsymbol{m} = \theta + \boldsymbol{m}_1 + \cdots + \boldsymbol{m}_d$ 的估计量定义为

$$\hat{\boldsymbol{m}} = \hat{\theta} + \hat{\boldsymbol{m}}_1 + \cdots + \hat{\boldsymbol{m}}_d,$$

其中 $\hat{\theta} = \bar{Y}$, $\hat{\boldsymbol{m}}_j$ 是 $\boldsymbol{m}_j$ 在观测点处的非参数估计, 它们是下列关于未知向量 $\boldsymbol{m}_1,\cdots,\boldsymbol{m}_d$ 的方程组的解:

$$\begin{pmatrix} I & S_1 & \cdots & S_1 \\ S_2 & I & \cdots & S_2 \\ \vdots & \vdots & & \vdots \\ S_d & S_d & \cdots & I \end{pmatrix} \begin{pmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \\ \vdots \\ \boldsymbol{m}_d \end{pmatrix} = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_d \end{pmatrix} \boldsymbol{Y}.$$

在实际应用中, 上述 nd 个未知量的 nd 个方程的方程组通常通过使用后移算法迭代来求解 (Buja et al., 1989), 但如果上述方程组的系数矩阵可逆, 那么可以将估计量写作

$$\begin{pmatrix} \hat{\boldsymbol{m}}_1 \\ \hat{\boldsymbol{m}}_2 \\ \vdots \\ \hat{\boldsymbol{m}}_d \end{pmatrix} = \begin{pmatrix} I & S_1 & \cdots & S_1 \\ S_2 & I & \cdots & S_2 \\ \vdots & \vdots & & \vdots \\ S_d & S_d & \cdots & I \end{pmatrix}^{-1} \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_d \end{pmatrix} \boldsymbol{Y}.$$

上述表达式说明可加模型的估计量也是线性光滑器. 可以将光滑矩阵 S_j 中心化以确保其有意义, 即用 $S_j^* = (I - \mathbf{1}\mathbf{1}^{\mathrm{T}})S_j$ 代替 S_j, $j=1,\cdots,d$.

Opsomer 和 Ruppert(1997) 与 Opsomer(2000) 证明了上述估计量的存在性. 他们给出了估计量的偏差和方差的渐近表示. 为表述这些结果, 首先引入下面的记号. 对任何 $j,k=1,\cdots,d$, 记

$$\mathcal{D}^p\boldsymbol{m}_j = \left(\frac{\mathrm{d}^p m_j(X_{1j})}{\mathrm{d}x_j},\cdots,\frac{\mathrm{d}^p m_j(X_{nj})}{\mathrm{d}x_j}\right)^{\mathrm{T}},$$

$$E[m_j^{(p)}(X_{ij})|\mathbf{X}_k] = \left(E[m_j^{(p)}(X_{1j})|X_{1k}],\cdots,E[m_j^{(p)}(X_{nj})|X_{nk}]\right),$$

设矩阵 T_{12}^* 的 (i,j) 元素为

$$[T_{ij}^*]_{ij}=\frac{1}{n}\frac{f(X_{i1},X_{i2})}{f_1(X_{i1})f_2(X_{i2})}-\frac{1}{n},$$

其中 $f(x_1,x_2)$ 表示 (X_{i1},X_{i2}) 的联合密度, $f_j(x_j)$ 表示 X_{ij} 的边际密度, $j=1,\cdots,d$. 对任何核函数 K, 其 r 阶矩记作 $\mu_r(K)=\int u^r K(u)\mathrm{d}u$, 并记 $R(K)=\int K^2(u)\mathrm{d}u$. 定义 K 关于 x 的边界矩为

$$\mu_r(K,x_j)=\int_{D_{x_j,h_j}} u^r K(u)\mathrm{d}u,\quad j=1,\cdots,d,$$

并记 $R(K,x_j)=\int_{D_{x_j,h_j}} K^2(u)\mathrm{d}u$, 其中

$$D_{x_j,h_j}=\{t:(x_j+h_jt)\in\mathrm{supp}(f_j)\}\cap\mathrm{supp}(K),\quad j=1,\cdots,d.$$

显然, 当 x_j 为内点时, $\mu_r(K,x_j)=\mu_r(K)$, $R(K,x_j)=R(K)$. 用 N_p 表示 $(p+1)\times(p+1)$ 矩阵, 其 (i,j) 元素等于 $\mu_{i+j-2}(K)$, $M_p(u)$ 为用 $(1,u,\cdots,u^p)^{\mathrm{T}}$ 代替 N_p 的第一列元素后得到的矩阵. 定义核 $K_{(p)}(u)=[|M_p(u)|/|N_p|]K(u)$. 对于边界矩, 定义矩阵 $N_p(x)$ 和 $M_p(u,x)$ 同上, 只是用 $\mu_r(K,x_j)$ 代替 $\mu_r(K)$ 即可. 我们也定义边界核

$$K_{(p)}(u,x_j)=[|M_p(u,x_j)|/N_p(x_j)|]K(u)I(u\in D_{x_j,h_j}),\quad j=1,\cdots,d.$$

如果 x_j 不在 $\mathrm{supp}(f)$ 的边界上, 那么 $K_{(p)}(u,x_j)=K_{(p)}(u)$. 用 $\boldsymbol{t}_i^{\mathrm{T}}$ 和 v_j 分别表示 $(I-T_{12}^*)^{-1}$ 的第 i 行和第 j 列, $\boldsymbol{e}_i^{\mathrm{T}}$ 表示第 i 个单位向量.

Opsomer 和 Ruppert(1997) 在 $d=2$ 和 p_1 和 p_2 是奇数的情况下给出了下列结果:

$$E(\hat{\theta}-\theta|\boldsymbol{X}_1,\boldsymbol{X}_2)=\theta+O_P\left(\frac{1}{\sqrt{n}}\right),$$

$$\mathrm{var}(\hat{\theta})=n^{-1}\sigma^2,$$

$$\begin{aligned}
&E[\hat{m}_1(X_{i1})-m_1(X_{i1})|\boldsymbol{X}_1,\boldsymbol{X}_2]\\
=&\frac{1}{(p_1+1)!}h_1^{p_1+1}\mu_{p_1+1}(K_{(p_1)},X_{i1})m_1^{(p_1+1)}(X_{i1})\\
&+\frac{1}{(p_1+1)!}h_1^{p_1+1}\mu_{p_1+1}(K_{(p_1)})\left\{(\boldsymbol{t}_i-\boldsymbol{e}_i)^{\mathrm{T}}\mathcal{D}^{p_1+1}\boldsymbol{m}_1-E[m_1^{(p_1+1)}(X_{i1})]\right\}\\
&-\frac{1}{(p_2+1)!}h_2^{p_2+1}\mu_{p_2+1}(K_{(p_2)})\left\{\boldsymbol{t}_i^{\mathrm{T}}E[m_2^{(p_2+1)}(X_{i2})|\boldsymbol{X}_1]-E[m_2^{(p_2+1)}(X_{i2})]\right\}\\
&+O_P(n^{-1/2})+o_P(h_1^{p_1+1}+h_2^{p_2+1}),
\end{aligned}$$

$$\mathrm{var}(\hat{m}_1(X_{i1})|\boldsymbol{X}_1, \boldsymbol{X}_2) = \frac{R(K_{(p_1)}, X_{i1})}{nh_1}\sigma^2 f_1^{-1}(X_{i1}) + o_P\left(\frac{1}{nh_1}\right).$$

同理, 可以得到 $\hat{m}_2$ 的条件偏差和方差的渐近表达式, 这只需将 X_{i1} 与 X_{i2} 互换即可. 由此可得到它们的平均均方误差的渐近表达式. 估计量 $\hat{m}(X_{i1}, X_{i2}) = \hat{\theta} + \hat{m}_1(X_{i1}) + \hat{m}_d(X_{i2})$ 的条件偏差和方差为

$$\begin{aligned}
&E[\hat{m}(X_{i1}, X_{i2}) - m(X_{i1}, X_{i2})|\boldsymbol{X}_1, \boldsymbol{X}_2]\\
=&\frac{1}{(p_1+1)!}h_1^{p_1+1}\Big\{\mu_{p_1+1}(K_{(p_1)}, X_{i1})m_1^{(p_1+1)}(X_{i1}) + \mu_{p_1+1}(K_{(p_1)})\\
&\times\left[(\boldsymbol{t}_i - \boldsymbol{e}_i)^{\mathrm{T}}\mathcal{D}^{p_1+1}\boldsymbol{m}_1 - v_i^{\mathrm{T}}E(m_1^{(p_1+1)}(X_{i1})|X_{i2})\right]\Big\}\\
&-\frac{1}{(p_2+1)!}h_2^{p_2+1}\Big\{\mu_{p_2+1}(K_{(p_2)}, X_{i2})m_2^{(p_2+1)}(X_{i2}) + \mu_{p_2+1}(K_{(p_2)})\\
&\times\left[(\boldsymbol{t}_i - \boldsymbol{e}_i)^{\mathrm{T}}\mathcal{D}^{p_2+1}\boldsymbol{m}_2 - v_i^{\mathrm{T}}E(m_2^{(p_2+1)}(X_{i2})|X_{i1})\right]\Big\}\\
&+o_P(h_1^{p_1+1} + h_2^{p_2+1}),
\end{aligned}$$

$$\begin{aligned}
&\mathrm{var}(\hat{m}(X_{i1}, X_{i2})|\boldsymbol{X}_1, \boldsymbol{X}_2)\\
=&\sigma^2\left(\frac{R(K_{(p_1)}, X_{i1})}{nh_1}f_1^{-1}(X_{i1}) + \frac{R(K_{(p_2)}, X_{i2})}{nh_2}f_2^{-1}(X_{i2})\right)\\
&+o_P\left(\frac{1}{nh_1} + \frac{1}{nh_2}\right),
\end{aligned}$$

当 p_1 和 p_2 为偶数时, Opsomer 和 Ruppert(1997) 也给出了估计量 $\hat{m}_1$、$\hat{m}_2$ 和 $\hat{m}$ 的条件偏差和方差的渐近表达式. 如果诸观测点 X_{ij} 相互独立, 那么估计量 $\hat{m}_j$ 在支撑集内部的观察点 X_{ij} 处的条件偏差和方差为

$$\begin{aligned}
&E[\hat{m}_j(X_{ij}) - m_j(X_{ij})|\boldsymbol{X}_1, \cdots, \boldsymbol{X}_d]\\
=&\frac{1}{(p_j+1)!}h_j^{p_j+1}\mu_{p_j+1}(K_{(p_j)})\left\{m_j^{(p_j+1)}(X_{ij}) - E[m_j^{(p_j+1)}(X_{ij})]\right\}\\
&+O_P(n^{-1/2}) + o_P(h_1^{p_1+1} + \cdots + h_d^{p_d+1}),
\end{aligned}$$

$$\mathrm{var}(\hat{m}_j(X_{ij})|\boldsymbol{X}_1, \cdots, \boldsymbol{X}_d) = \frac{R(K_{(p_j)})}{nh_j}\sigma^2 f_j^{-1}(X_{ij}) + o_P\left(\frac{1}{nh_j}\right).$$

估计量 $\hat{m}$ 在支撑集内部的观察点 X_{ij} 处的条件偏差为

$$\begin{aligned}
&E[\hat{m}(X_{i1}, \cdots, X_{id}) - m(X_{i1}, \cdots, X_{id})|\boldsymbol{X}_1, \cdots, \boldsymbol{X}_d]\\
=&\sum_{j=1}^{d}\frac{1}{(p_j+1)!}h_j^{p_j+1}\mu_{p_j+1}(K_{(p_j)})\left\{m_j^{(p_j+1)}(X_{ij}) - E[m_j^{(p_j+1)}(X_{ij})]\right\}\\
&+o_P(h_1^{p_1+1} + \cdots + h_d^{p_d+1}),
\end{aligned}$$

其中 $\hat{m}(X_{i1},\cdots,X_{id})=\hat{\theta}+\hat{m}_1(X_{i1})+\cdots+\hat{m}_d(X_{id})$. $O_P(n^{-1/2})$ 在 $\hat{m}$ 的渐近偏差中消失, 是因为 $\hat{\theta}$ 和 $\hat{m}_j(X_{ij})$ 的偏差项之间相互抵消了.

10.2.2 样条后移核估计

本节假定模型误差 ε_i 是异方差的, 即 $\mathrm{var}(\varepsilon_i|X_{i1},\cdots,X_{id})=\sigma^2(X_{i1},\cdots,X_{id})$, $i=1,\cdots,n$. 记 $\boldsymbol{Y}=(Y_1,\cdots,Y_n)^{\mathrm{T}}$, $X_i=(X_{i1},\cdots,X_{id})^{\mathrm{T}}$, $i=1,\cdots,n$, $x=(x_1,\cdots,x_d)$. 假定协变量 X_{ij} 在闭区间 $[a_j,b_j]$ 上取值, $j=1,\cdots,d$. 不失一般性, 不妨假定所有区间 $[a_j,b_j]=[0,1]$. 预先选择一个整数 $N_n\sim n^{2/5}$. 记 $H=H_n=(N+1)^{-1}$, $\{I_{kj}=[kH,(k+1)H),0\leqslant k\leqslant N\}$ 为 $N+1$ 个长度为 H 的 $[0,1]$ 的等距子区间. 在上 I_{kj} 上定义 $N+1$ 个示性函数 $I_{k,j}(x_j)=I_{I_{kj}}(x_j)$, $k=0,1,\cdots,N$.

定义可加样条函数的 $1+dN$ 维空间 G 为由 $\{1,I_{k,j}(x_j),1\leqslant j\leqslant d,1\leqslant k\leqslant N\}$ 张成的线性空间, 并用 $G_n\subset\mathbf{R}^n$ 表示由 $\{1,[I_{k,j}(X_{ij})]_{i=1}^n,1\leqslant j\leqslant d,1\leqslant k\leqslant N\}$ 张成的线性空间. 当 $n\to\infty$, G_n 的维数依概率趋近于 1 变为 $1+dN$. 可加函数 $m(x)$ 的样条估计量是来自空间 G 的唯一元素 $\hat{m}(x)$, 以致向量 $(\hat{m}(X_1),\cdots,\hat{m}(X_n))^{\mathrm{T}}$ 能最佳逼近响应向量 $\boldsymbol{Y}$. 为明确起见, 定义

$$\hat{m}(x)=\hat{\lambda}_0+\sum_{j=1}^{d}\sum_{k=1}^{N}\hat{\lambda}_{k,j}I_{k,j}(x_j), \tag{10.2.2}$$

其中系数 $(\hat{\lambda}_0,\hat{\lambda}_{1,1},\cdots,\hat{\lambda}_{N,d})$ 是下列最小二乘问题的解:

$$(\hat{\lambda}_0,\hat{\lambda}_{1,1},\cdots,\hat{\lambda}_{N,d})^{\mathrm{T}}=\arg\min_{\mathbf{R}^{dN+1}}\sum_{i=1}^{n}\left\{Y_i-\lambda_0-\sum_{j=1}^{d}\sum_{k=1}^{N}\lambda_{k,j}I_{k,j}(X_{ij})\right\}^2.$$

分量函数 $m_j(x_j)$ 和常数 θ 的引导估计量为

$$\hat{m}_j(x_j)=\sum_{k=1}^{N}\hat{\lambda}_{k,j}I_{k,j}(x_j)-\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{N}\hat{\lambda}_{k,j}I_{k,j}(X_{ij}), \tag{10.2.3}$$

$$\hat{m}_\theta=\hat{\lambda}_0+\frac{1}{n}\sum_{j=1}^{d}\sum_{i=1}^{n}\sum_{k=1}^{N}\hat{\lambda}_{k,j}I_{k,j}(X_{ij}).$$

然后, 这些引导估计量用来定义一个新伪响应集 $\hat{Y}_{i1}$, 用它作为不可观测 "原" 响应 Y_{i1}, 即

$$\hat{Y}_{i1}=Y_i-\hat{\theta}-\sum_{j=2}^{d}\hat{m}_j(X_{ij}),\quad Y_{i1}=Y_i-\theta-\sum_{j=2}^{d}m_j(X_{ij}), \tag{10.2.4}$$

其中 $\hat{\theta}=n^{-1}\sum_{i=1}^{n}Y_i$. 由中心极限定理可知 $\hat{\theta}$ 是 θ 的 $\sqrt{n}$ 相合估计量. 基于 $\{(X_{i1},\hat{Y}_{i1}),1\leqslant i\leqslant n\}$, 定义 $m_1(x_1)$ 的样条后移核 (SBK) 估计量为 $\hat{m}_1^*(x_1)$, 它试图模仿基于 $\{(X_{i1},Y_{i1}),1\leqslant i\leqslant n\}$ 的 $m_1(x_1)$ 的准 Nadaraya-Watson 估计量 $\tilde{m}_1^*(x_1)$, 其中不可观测 "原" 响应 $\{Y_{i1},1\leqslant i\leqslant n\}$ 是可利用的. 亦即

$$\hat{m}_1^*(x_1)=\frac{\sum_{i=1}^{n}K_h(X_{i1}-x_1)\hat{Y}_{i1}}{\sum_{i=1}^{n}K_h(X_{i1}-x_1)},$$

$$\tilde{m}_1^*(x_1)=\frac{\sum_{i=1}^{n}K_h(X_{i1}-x_1)Y_{i1}}{\sum_{i=1}^{n}K_h(X_{i1}-x_1)},$$

其中 $\hat{Y}_{i1}$ 和 Y_{i1} 在 (10.2.4) 式中定义, $K_h(\cdot)=h^{-1}K(\cdot/h)$, $K(\cdot)$ 是某个核函数, $h=h_n$ 是带宽.

类似于 $\hat{m}_1^*(x_1)$ 的构造, 可以定义 $\hat{m}_j^*(x_j)$, $j=2,\cdots,d$, 即

$$\hat{m}_j^*(x_j)=\frac{\sum_{i=1}^{n}K_h(X_{ij}-x_j)\hat{Y}_{ij}}{\sum_{i=1}^{n}K_h(X_{ij}-x_j)},$$

$$\hat{Y}_{ij}=Y_i-\hat{\theta}-\sum_{1\leqslant l\leqslant d,l\neq j}\hat{m}_l(X_{il}),$$

其中 $\hat{m}_l(X_{il})$, $l=1,\cdots,d$ 是由 (10.2.3) 式定义的每个分量函数的引领估计. 用局部多项式估计代替 Nadaraya-Watson 估计量也可以构造类似的估计量.

Wang 和 Yang(2007) 与 Wang 和 Yang(2009) 在一定的条件下证明了 $\hat{m}_j^*(x_j)$ 的渐近正态性, 并给出了它的一致收敛速度. 具体地说, 有下列结果: 在一些正则条件下, 对任何 $j=1,\cdots,d$, 有

$$\sup_{x_j\in[0,1]}\left|\hat{m}_j^*(x_j)-\tilde{m}_j^*(x_j)\right|=o_P(n^{-2/5}),$$

且对任何 $x_j\in[h,1-h]$, 有

$$\sqrt{nh}[\hat{m}_j^*(x_j)-m_j(x_j)-b_j(x_j)h^2]\xrightarrow{D}N(0,v_j^2(x_j)),$$

其中

$$b_j(x_j) = \int u^2K(u)\mathrm{d}u[m_j''(x_j)f_j(x_j)/2 + m_j'(x_j)f_j'(x_j)]f_j^{-1}(x_j),$$

$$v_j^2(x_j) = \int K^2(u)\mathrm{d}uE[\sigma^2(X_{11},\cdots,X_{1d}|X_{1j}=x_j]f_j^{-1}(x_j).$$

设

$$\hat{m}(x) = \hat{\theta} + \sum_{j=1}^{d}\hat{m}_j^*(x_j), \quad b(x) = \sum_{j=1}^{d}b_j(x_j), \quad v^2(x) = \sum_{j=1}^{d}v_j^2(x_j),$$

则

$$\sqrt{nh}[\hat{m}^*(x) - m(x) - b(x)h^2] \xrightarrow{D} N(0, v^2(x)).$$

上述结果可以构造 $\hat{m}_j(x_j)$ 和 $\hat{m}(x)$ 的近似逐点置信区间.

10.2.3 多项式样条估计

1. **估计**

本节使用多项式样条光滑方法估计模型 (10.2.1) 中的非参数函数. 它涉及非参数函数 $\{m_j(\cdot), 1 \leqslant j \leqslant d\}$ 的多项式样条逼近.

选定一个正整数 $N = N_n$. 对每个 $1 \leqslant j \leqslant d$, 引入 $[0,1]$ 上具有 N 个内节点的节点序列 k_j, 即 $k_j = \{0 = v_{0,j} < v_{1,j} < \cdots < v_{N,j} < v_{N+1,j} = 1\}$. 记 $I_{lj} = [v_{l,j}, v_{l+1,j})$, $l = 0, 1, \cdots, N-1$, $I_{Nj} = [v_{N,j}, v_{N+1,j}]$. 特别地, 可以将 $[0,1]$ 进行等距分割, 即 $v_{l,j} = l/N$, $l = 0, 1, \cdots, N-1$. 对任何非负整数 p, 用 $\mathcal{G}_j = \mathcal{G}^p([0,1], k_j)$ 表示 $[0,1]$ 上的函数 g 构成的空间, 使得

(i) g 是每个区间 $I_{lj}(0 \leqslant l \leqslant N)$ 上的 p 阶多项式;

(ii) 对 $p \geqslant 1$, g 是 $[0,1]$ 上 $p-1$ 阶连续可微函数.

称满足 (i) 和 (ii) 的函数为多项式样条. 分段多项式把内节点光滑地连接起来. 例如, 阶 $p=0$ 的多项式样条是分段常数函数, 阶 $p=1$ 的多项式样条是 $[0,1]$ 上连续的分段线性函数. 多项式样条空间 $\mathcal{G}_j$ 由多项式的阶数 p 和节点序列 k_j 所确定. 设 $h_j = h_{n,j} = \max\limits_{0 \leqslant l \leqslant N}|v_{l+1,j} - v_{l,j}|$; h_j 正像局部多项式光滑中的带宽一样可以作为一个光滑参数做欠光滑. 定义 $h = \max\limits_{1 \leqslant j \leqslant p} h_j$; 用 h 度量过光滑.

为了相合地估计 (10.2.1) 式中具有 $E[m_j(X_{ij})] = 0$ 的中心化函数 $\{m_j(x_j), 1 \leqslant j \leqslant d\}$, 引入经验中心化多项式样条

$$\mathcal{G}_j^0 = \left\{g_j : g_j \in \mathcal{G}_j, \frac{1}{n}\sum_{i=1}^{n}g_j(X_{ij}) = 0\right\}.$$

可以很方便地构造 $\mathcal{G}_j^0$ 的基. 例如, 在实施中使用经验中心化截断幂基, 即

$$\left\{B_{rj} = b_{rj} - \frac{1}{n}\sum_{i=1}^{n} b_{rj}(X_{ij}), 1 \leqslant r \leqslant R\right\},$$

其中 $R \equiv R_n = N_n + p$, 而 $\{b_{1j}, \cdots, b_{Rj}\}$ 是下面给出的截断幂基:

$$\{x_j, \cdots, x_j^p, (x_j - v_{1,j})_+^p, \cdots, (x_j - v_{N_j})_+^p\}.$$

此处 $(x)_+^p = (x_+)^p$. 如果函数 $\{m_j(x_j), 1 \leqslant j \leqslant d\}$ 足够光滑, 那么可以用多项式样条 $\{g_j(x_j) \in \mathcal{G}_j^0, 1 \leqslant j \leqslant d\}$ 逼近他们. 即对每个 $1 \leqslant j \leqslant d$, 利用系数集 $\{\gamma_{rj}, 1 \leqslant r \leqslant R\}$, 有

$$m_j(x_j) \approx g_j(x_j) = \sum_{r=1}^{R} \gamma_{rj} B_{rj}(x_j). \tag{10.2.5}$$

定义目标函数为

$$Q(\theta, \gamma) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \theta - \sum_{j=1}^{d}\sum_{r=1}^{R} \gamma_{rj} B_{rj}(X_{ij})\right]^2, \tag{10.2.6}$$

其中系数 $\gamma = \{\gamma_{rj}, 1 \leqslant r \leqslant R, 1 \leqslant j \leqslant d\}$. 极小化 $Q(\theta, \gamma)$ 可以的到它的极小值点 $(\hat{\theta}, \hat{\gamma})$, 即

$$(\hat{\theta}, \hat{\gamma}) = \arg\max_{\theta,\gamma} Q(\theta, \gamma).$$

因此, 模型 (10.2.1) 中可加函数的估计量为

$$\hat{m}_j(x_j) = \sum_{r=1}^{R} \hat{\gamma}_{rj} B_{rj}(x_j), \quad j = 1, \cdots, d.$$

$m(x) = \theta + m_1(x_1) + \cdots + m_d(x_d)$ 的估计量为

$$\hat{m}(x) = \hat{\theta} + \sum_{j=1}^{d} \hat{m}_j(x_j). \tag{10.2.7}$$

为了给出了 $\hat{m}_j(x_j)$ 和 $\hat{m}(x)$ 的收敛速度, 我们引入下列记号. 用 $C^p([0,1])$ 表示 p 次连续可微的函数空间. 记 $\mathcal{X} = [0,1]^d$. 设 $X_i = (X_{i1}, \cdots, X_{id})^{\mathrm{T}} \in \mathcal{X}$, $i = 1, \cdots, n$, X 与 X_i 同分布. 又设 $E(Y^2) < \infty$, 那么 Y 关于 X 的回归函数为 $m(x) = E(Y|X = x) = \theta + \sum_{j=1}^{d} m_j(x_j)$, 条件方差为 $\sigma^2(x) = \mathrm{var}(Y|X = x)$, $x \in \mathcal{X}$. 用 $||\psi||$ 表示函数 ψ 的 L^2 模, 即 $||\psi|| = \{E[\psi^2(X)]\}^{1/2}$.

类似于 (Stone, 1985) 的推论 1 和 (Xue and Yang, 2006a) 的定理 1, 可以得到下列结果.

定理 10.2.1 设 X 的联合密度 $f(x)$ 在支撑 $\mathcal{X}$ 上绝对连续, 且存在正的常数 c_1 和 c_2 使得 $c_1 \leqslant \inf_{x\in\mathcal{X}} f(x) \leqslant \sup_{x\in\mathcal{X}} f(x) \leqslant c_2$; 方差函数 $\sigma^2(x)$ 在 $\mathcal{X}$ 上可测且有界; $m_j \in C^p([0,1])$, $1 \leqslant j \leqslant d$; 对节点集合 k_j, 存在常数 $c > 0$, 使得

$$\max_{1\leqslant j\leqslant d} \frac{\max\{v_{l+1,j} - v_{l,j}, l = 0, \cdots, N\}}{\min\{v_{l+1,j} - v_{l,j}, l = 0, \cdots, N\}} \leqslant c;$$

如果内节点数 $N_n \asymp n^{1/(2p+3)}$, 则

$$E(\hat{\theta} - \theta)^2) = O_P(n^{-1}),$$

$$\max_{1\leqslant j\leqslant d} ||\hat{m}_j - m_j|| = O_P(h^{p+1} + (nh)^{-1/2}),$$

$$E(||\hat{m} - m|| = O_P(h^{p+1} + (nh)^{-1/2}),$$

其中 p 是样条的阶, $\asymp$ 表示两边有相同的阶.

如果向量过程 $\{\xi_t\}_{t=-\infty}^{\infty} = \{(X_t, Y_t)\}_{t=-\infty}^{\infty}$ 是严平稳和几何强混合的, 其 α 混合系数 $\alpha(k) \leqslant c\rho^k$, $c > 0$ 和 $0 < \rho < 1$ 为常数, 则定理 10.2.1 的结果亦成立, 其中 $\alpha(k) = \sup_{A\in\sigma(\xi_t, t\leqslant 0), B\in\sigma(\xi_t, t\geqslant k)} |P(A)P(B) - P(AB)|$. 参看 Xue 和 Yang(2006a).

定理 10.2.1 的收敛速度不依赖于 d. h 的最优阶为 $n^{-1/(2p+3)}$, 在这种情况下, $||\hat{m}_j - m_j|| = O_P(n^{-1/(2p+3)})$. 从 Stone(1982) 对 $d = 1$ 给出的收敛速度可以看出, 定理 10.2.1 的收敛速度在非参数估计的意义下是最优的. 因此, 可加模型有效地避免了维数灾祸.

2. 节点数选择

多项式样条估计方法取决于节点序列 $\{k_j, 1 \leqslant j \leqslant d\}$ 的适当选择, 尤其内节点数 N_n. 因此, 适当选择节点序列对有效地改进估计是非常重要的. 这里, 我们使用 AIC 准则选择 N. 对节点位置, 我们使用等距节点或分位数节点 (样本分位数与任意两个相邻节点之间的观察数目相同).

根据定理 10.2.1, N_n 的最优阶为 $n^{1/(2p+3)}$. 因此, 我们建议从 $[0.5N_r, \min(5N_r, Tb)]$ 中选择最优节点数 $\hat{N}_n^{\text{opt}}$ 以确保最小二乘估计中参数的总数小于 $n/4$, 其中 $N_r = n^{1/(2p+3)}$, $Tb = (n/4 - 1)/d$.

用 $\hat{Y}_i(N_n) = \hat{m}(X_i)$ 表示第 i 个响应 Y_i 的估计量, $i = 1, \cdots, n$. 这里给定在 (10.2.7) 式中的 $\hat{m}$ 依赖于节点序列. 设 $q_n = dR_n + 1$ 是 (10.2.6) 式中参数的总数, 那么 $\hat{N}_n^{\text{opt}}$ 是下列 AIC 函数的极小值点:

$$\hat{N}_n^{\text{opt}} = \arg \min_{N_n\in[0.5N_r, \min(5N_r, Tb)]} \text{AIC}(N_n),$$

其中 $\text{AIC}(N_n) = \log(\text{MSE}) + 2q_n/n$, MSE$=n^{-1}\sum_{i=1}^{n}[Y_i - \hat{Y}_i(N_n)]^2$.

10.3 变量选择

本节假定模型 (10.2.1) 的误差 ε_i 的方差为 σ^2. 我们允许 d 大于样本量 n, 且 d 随着 n 的增大而增大. 本节使用自适应组 LASSO 来选择模型 (10.2.1) 中的变量, 并证明提出的方法能够以很高的概率选择非零分量. 自适应组 LASSO 是将自适应 LASSO(Zou, 2006) 简单推广到组 LASSO(Yuan and Lin, 2006) 的方法. LASSO 方法由 Tibshirani(1996) 首次提出, 它的英文全称是 "least absolute shrinkage and selection operator". 基于可加模型的每一个分量的样条表示, 我们描述使用组 LASSO 进行变量选择的两步方法. 第一步使用组 LASSO 来减少模型维数, 并得到非参数分量的初始估计量. 第二步使用自适应组 LASSO 来选择非参数分量的最终集. 本节的内容主要取材于文献 (Huang et al., 2010).

假定非零 m_j 的数目是固定的, 这能够使我们在简单假定下得到模型选择的相合性. 本节仍使用 10.2.3 小节中的记号. 下面描述的变量选择方法是基于表达式 (10.2.5).

记 $\gamma_j=(\gamma_{1j},\cdots,\gamma_{Rj})^{\mathrm T}$, $\gamma=(\gamma_1^{\mathrm T},\cdots,\gamma_d^{\mathrm T})^{\mathrm T}$. $w=(w_1,\cdots,w_d)^{\mathrm T}$ 是给定的权向量, 其中 $0\leqslant w_j\leqslant\infty$, $1\leqslant j\leqslant d$. 考虑惩罚最小二乘准则

$$L(\theta,\gamma)=\sum_{i=1}^{n}\left[Y_i-\theta-\sum_{j=1}^{d}\sum_{r=1}^{R}\gamma_{rj}B_{rj}(X_{ij})\right]^2+\lambda\sum_{j=1}^{d}w_j\|\gamma_j\|, \tag{10.3.1}$$

其中 λ 是一个惩罚参数, $\|\cdot\|$ 表示 Euclid 模. 通过极小化 $L(\theta,\gamma)$ 可以得到所需要的估计量. 为了更方便地描述两步估计, 下面我们将 (10.3.1) 式重新表示. 记 $Z_{ij}=(B_{1j}(X_{ij}),\cdots,B_{Rj}(X_{ij}))^{\mathrm T}$. 因此 Z_{ij} 是由第 j 个协变量的第 i 个观察处 (中心化) 基函数的值组成. 设 $\boldsymbol{Z}_j=(Z_{1j},\cdots,Z_{nj})^{\mathrm T}$ 为相应于第 j 个协变量的 $n\times\boldsymbol{R}$"设计" 矩阵. 记 $\boldsymbol{Y}^*=(Y_1-\bar Y,\cdots,Y_n-\bar Y)^{\mathrm T}$. 利用这些记号, 可以将 (10.3.1) 式重新写为

$$L(\gamma,\lambda)=\|\boldsymbol{Y}^*-\boldsymbol{Z}\gamma\|^2+\lambda\sum_{j=1}^{d}w_j\|\gamma_j\|. \tag{10.3.2}$$

这里我们在 L 中去掉了 θ, 因为 θ 的估计为 $\bar Y$. 因此, 极小化 (10.3.1) 式等价于关于 γ 极小化 (10.3.2) 式.

下面描述可加模型 (10.2.1) 中分量选择的两步方法.

步骤 1. 计算组 LASSO 估计量. 设

$$L_1(\gamma,\lambda_1)=\|\boldsymbol{Y}^*-\boldsymbol{Z}\gamma\|^2+\lambda_1\sum_{j=1}^{d}\|\gamma_j\|.$$

这个目标函数是 (10.3.2) 式的特殊情况, 它是通过设置 $w_j = 1(1 \leqslant j \leqslant d)$ 而得到. 组 LASSO 估计量为 $\tilde{\gamma} \equiv \tilde{\gamma}(\lambda_1) = \arg\min_{\gamma} L_1(\gamma, \lambda_1)$.

步骤 2. 使用组 LASSO 估计量 $\tilde{\gamma}$ 得到权 w_j, 即

$$w_j = \begin{cases} ||\tilde{\gamma}_j||^{-1}, & ||\tilde{\gamma}_j|| > 0 \\ \infty, & ||\tilde{\gamma}_j|| = 0. \end{cases}$$

自适应 LASSO 的目标函数是

$$L_2(\gamma, \lambda_2) = ||\boldsymbol{Y}^* - \boldsymbol{Z}\gamma||^2 + \lambda_2 \sum_{j=1}^{d} ||\gamma_j||,$$

这里定义 $0 \cdot \infty = 0$. 因此, 组 LASSO 没有选出的分量不包括在步骤 2 中. 自适应组 LASSO 估计量为 $\hat{\gamma} \equiv \hat{\gamma}(\lambda_2) = \arg\min_{\gamma} L_2(\gamma, \lambda_2)$. 最后 θ 和 m_j 的自适应组 LASSO 估计量为

$$\hat{\theta} = \bar{Y} \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \hat{m}_j(x_j) = \sum_{r=1}^{R_n} \hat{\gamma}_{rj} B_{rj}(x_j), \quad j = 1, \cdots, d.$$

下面考虑自适应组 LASSO 的选择相合性. 为此, 需要定义下列光滑函数空间. 设 k 是一个非负整数, $\delta \in (0,1]$, 使得 $p = k + \delta > 0.5$. 用 $\mathcal{M}$ 表示 $[0,1]$ 上的函数构成的集合, 其中每一个元素 φ 具有 k 阶导数, 且满足 δ 阶 Hölder 条件:

$$|\varphi^{(k)}(t') - \varphi^{(k)}(t)| \leqslant C|t' - t|^{\delta}, \quad 0 \leqslant t, t' \leqslant 1,$$

其中 $C > 0$ 为常数.

不失一般性, 在模型 (10.2.1) 中, 假定前 q 个分量非零, 即 $m_j(x_j) \neq 0$, $1 \leqslant j \leqslant q$, 但 $m_j(x_j) \equiv 0$, $q+1 \leqslant j \leqslant d$. 记 $A_1 = \{1, \cdots, q\}$, $A_0 = \{q+1, \cdots, d\}$. 对任何函数 φ, 定义 $||\varphi||_2 = \left[\int_0^1 \varphi^2(t)\mathrm{d}t\right]^{1/2}$.

Huang 等 (2010) 证明了下列结果.

定理 10.3.1 设 X 的联合密度 $f(x)$ 在支撑 $\mathcal{X} = [0,1]^d$ 上连续, 且存在正的常数 c_1 和 c_2 使得 $c_1 \leqslant \inf_{x\in\mathcal{X}} f(x) \leqslant \sup_{x\in\mathcal{X}} f(x) \leqslant c_2$; 非零分量数 q 是固定的, $m_j \in \mathcal{M}$, $1 \leqslant j \leqslant q$, 且存在常数 $c_3 > 0$ 使得 $\min_{1\leqslant j\leqslant q} ||m_j||_2 \geqslant c_3$; 对所有 $t \geqslant 0$ 和对正的常数 c_4 和 c_5, 有 $P(|\varepsilon_i| > t) \leqslant c_4 \exp(-c_5 t^2)$, $i = 1, \cdots, n$. 又设组 LASSO 估计量 $\tilde{\gamma} = \tilde{\gamma}(\lambda_1)$ 是自适应组 LASSO 的初始估计量, 且具有 $\lambda_1 \asymp [n\log(dR)]^{1/2}$, $R \asymp n^{1/(2p+1)}$. 如果 $\lambda_2 \leqslant O(n^{1/2})$ 且满足

$$\lambda_2 / n^{(8p+3)/(8p+4)} = o(1), \quad n^{1/(4p+2)} \log^{1/2}(dR)/\lambda_2 = o(1), \tag{10.3.3}$$

则

$$P\left(\|\hat{m}_j\|_2 > 0, j \in A_1 \text{ 且 } \|\hat{m}_j\|_2 = 0, j \in A_0\right) \to 1,$$

$$\sum_{j=1}^{q} \|\hat{m}_j - m_j\|_2 = O_P\left(n^{-p/(2p+1)}\right).$$

从定理 10.3.1 可以看出, 允许的最大 λ_2 是 $\lambda_2 = O(n^{1/2})$. 利用这个 λ_2, 容易验证 (10.3.3) 的第一式满足. 将它代入 (10.3.3) 的第二式可得 $d = \exp\{o(n^{2p/(2p+1)})\}$, 它是允许的最大维数 d, 且可以大于 n. 因此, 在定理 10.3.1 的条件下, 使用组 LASSO 作为初始估计的自适应组 LASSO 估计量是选择相合的, 且达到了最优收敛速度, 即使当维数 d 大于 n 时也是如此. 遵照模型选择的思想, 使用现有方法可以得到非零分量的 oracle 有效的渐近正态估计量.

10.4 可加部分线性模型

将多元线性模型和可加模型进行叠加即可得到可加部分线性模型, 其形式为

$$Y_i = \beta_0^{\mathrm{T}} X_i + m_1(Z_{i1}) + \cdots + m_d(Z_{id}) + \varepsilon_i, \tag{10.4.1}$$

其中 β_0 是 p 维未知参数, 诸 m_j 是未知函数, 误差 $\{\varepsilon_i, 1 \leqslant i \leqslant n\}$ 是均值为 0 的 iid 随机变量, $X_i = (X_{i1}, \cdots, X_{ip})^{\mathrm{T}}$, Z_{ij} 为 Z_i 的第 j 个分量. 假设 $\{(X_i, Z_i, Y_i), 1 \leqslant i \leqslant n\}$ 为 iid 随机变量. 为记号简单, 我们在模型 (10.4.1) 中考虑 $d = 2$ 的情况. 为确保函数 m_1 和 m_2 的可识别性, 假设 $E[m_1(Z_{11})] = E[m_2(Z_{12})] = 0$. 记 $\boldsymbol{X} = (X_1, \cdots, X_n)^{\mathrm{T}}$, $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^{\mathrm{T}}$, $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^{\mathrm{T}}$, $\boldsymbol{m}_j = (m_j(Z_{1j}), \cdots, m_{nj}(Z_{nj}))^{\mathrm{T}}$, $j = 1, 2$. 在矩阵记号下, 可以将模型 (10.4.1) 表示为

$$\boldsymbol{Y} = \boldsymbol{X}\beta_0 + \boldsymbol{m}_1 + \boldsymbol{m}_2 + \boldsymbol{\varepsilon}. \tag{10.4.2}$$

本节主要阐述模型 (10.4.1) 中兴趣参数的估计方法、经验似然和变量选择问题. 本节内容主要取材包括: 10.4.1 小节取材于 (Opsomer and Ruppert, 1999; Liang et al., 2008); 10.4.2 小节取材于 (Liang et al., 2009; Wang et al., 2010); 10.4.3 小节取材于 (Zhou et al., 2011).

10.4.1 估计方法

下面利用后移算法对模型 (10.4.2) 中的参数和非参数分量进行估计. 我们仅考虑局部线性回归作为光滑方法. 可以直接扩展到高阶局部多项式估计, 其基本思想可参见 10.2.1 小节.

设 s^{T}_{1,z_1} 和 s^{T}_{2,z_2} 为在 z_1 和 z_2 处的局部线性回归的等价核, 即

$$s^{\mathrm{T}}_{j,z_j} = e^{\mathrm{T}}_j(\breve{Z}^{\mathrm{T}}_j W_j \breve{Z}_j)^{-1}\breve{Z}^{\mathrm{T}}_j W_j, \quad j=1,2,$$

其中 $e_1=(1,0)^{\mathrm{T}}$, $e_2=(0,1)^{\mathrm{T}}$,

$$W_j = \mathrm{diag}\left\{K_{h_j}(Z_{1j}-z_j),\cdots,K_{h_j}(Z_{nj}-z_j)\right\}, \quad j=1,2,$$

$K_{h_j}(\cdot)=h_j^{-1}K(\cdot/h_j)$, $K(\cdot)$ 为某个核函数, h_1 和 h_2 为带宽,

$$\breve{Z}_j = \begin{pmatrix} 1 & (Z_{1j}-z_j) \\ \vdots & \vdots \\ 1 & (Z_{nj}-z_j) \end{pmatrix}, \quad j=1,2.$$

设 S_j 为 m_j 在观测 $\boldsymbol{Z}_j$ 处局部多项式拟合的光滑矩阵, 即

$$S_j = (s_{j,Z_{1j}},\cdots,s_{j,Z_{nj}})^{\mathrm{T}}, \quad j=1,2.$$

那么 $S^*_j=(I-\mathbf{1}\mathbf{1}^{\mathrm{T}})S_j$ 是相应于 S_j 的中心化光滑矩阵, $j=1,2$. 可加部分线性模型的后移算法定义为下列方程组的解:

$$\begin{cases} \hat{\beta}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{Y}-\hat{m}_1-\hat{m}_2), \\ \hat{m}_1=S^*_1(\boldsymbol{Y}-\boldsymbol{X}\hat{\beta}-\hat{m}_2), \\ \hat{m}_2=S^*_2(\boldsymbol{Y}-\boldsymbol{X}\hat{\beta}-\hat{m}_1), \end{cases} \tag{10.4.3}$$

其中 $\hat{\boldsymbol{m}}_j=(\hat{m}_j(Z_{1j}),\cdots,\hat{m}_j(Z_{nj}))^{\mathrm{T}}$, $j=1,2$. 作为一个结果, β_0 的非迭代估计量具有形式

$$\hat{\beta}_{\mathrm{NE}} = [\boldsymbol{X}^{\mathrm{T}}(I-S)\boldsymbol{X}]^{-1}\boldsymbol{X}^{\mathrm{T}}(I-S)\boldsymbol{Y},$$

其中

$$S=[I-(I-S^*_1S^*_2)^{-1}(I-S^*_1)]+[I-(I-S^*_2S^*_1)^{-1}(I-S^*_2)]. \tag{10.4.4}$$

为确保 $\hat{\beta}_{\mathrm{NE}}$ 是 $\sqrt{n}$ 相合的, 需要作欠光滑. 我们将通过 profile 似然方法去掉这一限制 (Severini and Staniswalis, 1994). 基本思想描述如下. 对每个 z_1, z_2 和 β, 设 $\hat{m}_1(\beta,z_1)$ 和 $\hat{m}_2(\beta,z_2)$ 为 $m_1(z_1)$ 和 $m_2(z_2)$ 的后移估计量, 他们是用 β 代替 (10.4.3) 的后两式中的 $\hat{\beta}$ 后而得到. 相应地, $\hat{m}_1$ 和 $\hat{m}_2$ 可表示为

$$\begin{cases} \hat{m}_1(\beta)=[I-(I-S^*_1S^*_2)^{-1}(I-S^*_1)](Y-\beta^{\mathrm{T}}X), \\ \hat{m}_2(\beta)=[I-(I-S^*_2S^*_1)^{-1}(I-S^*_2)](Y-\beta^{\mathrm{T}}X). \end{cases} \tag{10.4.5}$$

将 $\hat{m}_1(\beta)$ 和 $\hat{m}_2(\beta)$ 代入 (10.4.3) 式并使用最小二乘法, 可得 β_0 的 profile 估计量:

$$\hat{\beta}_{\rm PE} = [\boldsymbol{X}^{\rm T}(I-S)^{\star 2}\boldsymbol{X}]^{-1}\boldsymbol{X}^{\rm T}(I-S)^{\star 2}\boldsymbol{Y}, \tag{10.4.6}$$

其中 $A^{\star 2} = A^{\rm T}A$.

正如 Opsomer 和 Ruppert(1997) 讨论的那样, 如果假设 $\sum_{i=1}^{n} m_1(Z_{i1}) = \sum_{i=1}^{n} m_2(Z_{i2}) = 0$, 那么为确保算法收敛, 中心化 S_1^* 和 S_2^* 是必需的, 且 (10.4.3) 式的估计量是很好的定义形式. $\hat{\beta}_{\rm PE}$ 和 $\hat{\beta}_{\rm NE}$ 之间的差别是出现在 $\hat{\beta}_{\rm PE}$ 中的因子为 $(I-S)^{\star 2}$, 这就避免了可加线性模型的欠光滑约束. 因为 $\hat{\beta}_{\rm PE}$ 的偏差由 $\boldsymbol{X}^{\rm T}(I-S)^{\star 2}(m_1+m_2)$ 确定, 它的每一个元素的阶为 $O(h_1^4+h_2^4)$, 利用通常的最优带宽 $n^{-1/5}$ 可推出 $\hat{\beta}_{\rm PE}$ 是 $\sqrt{n}$ 相合的.

Opsomer 和 Ruppert(1999) 在非参数函数估计的欠光滑下得到了 $\hat{\beta}_{\rm NE}$ 的 $\sqrt{n}$ 相合性, 他们要求的带宽为 $h_j \propto n^r$, $j=1,2$, $-1<r<-1/4$. Liang 等 (2008) 在不需要欠光滑下对 X 具有测量误差情形证明了 $\hat{\beta}_{\rm PE}$ 的 $\sqrt{n}$ 相合性, 他们假定的带宽为 $h_j \propto n^{-1/5}$, $j=1,2$, 这就放宽了对带宽的要求. 为描述 $\hat{\beta}_{\rm PE}$ 的渐近正态性, 需要给出下面一组条件.

(C10.4.1) $E(\varepsilon|X_1,Z_1)=0$, $\sup_{x,z} E(|\varepsilon|^3|X_1=x,Z_1=z)<\infty$, $\sup_{z} E(||X_1||^3|Z_1=z)<\infty$; 矩阵 $\Gamma=\mathrm{cov}(X_1^*)$ 为正定矩阵, 其中 $X_1^*=X_1-E(X_1|Z_{11})-E(X_2|Z_{12})$.

(C10.4.2) 带宽 h_1 和 h_2 的阶为 $n^{-1/5}$.

(C10.4.3) 核函数 $K(\cdot)$ 是具有紧支撑的有界对称的密度函数, 且 $\int u^2K(u)\mathrm{d}u=1$.

(C10.4.4) Z_{11} 和 Z_{12} 的密度函数 $f_1(z_1)$ 和 $f_2(z_2)$ 具有紧支撑并在其紧支撑集上有界远离 0, 且具有有界连续的二阶导数.

类似于 Liang 等 (2008) 的定理 1, 我们给出 $\hat{\beta}_{\rm PE}$ 的渐近分布, 即有下述结果:

定理 10.4.1　设条件 (C10.4.1)~(C10.4.4) 成立, 则

$$\sqrt{n}(\hat{\beta}_{\rm PE}-\beta_0) \xrightarrow{D} N(0,\Gamma^{-1}\Sigma\Gamma^{-1}),$$

其中 $\Sigma = E(\varepsilon^2 X_1^* X_1^{*T})$.

估计量 $\hat{\beta}_{\rm PE}$ 的渐近方差矩阵 $\Omega=\Gamma^{-1}\Sigma\Gamma$ 可以用 Sandwich 方法估计. 首先, 定义 Γ 和 Σ 的估计量为

$$\hat{\Gamma} = n^{-1}\boldsymbol{X}^{\rm T}(I-S)^{\star 2}\boldsymbol{X}, \quad \hat{\Sigma} = \frac{1}{n}\sum_{i-1}^{n}\hat{\varepsilon}_i^2\widetilde{X}_i\widetilde{X}_i^{\rm T},$$

其中 $\widetilde{X}_i$ 是 $(I-S)\boldsymbol{X}$ 的第 i 行. 于是, Ω 的估计量为 $\hat{\Omega}=\hat{\Gamma}^{-1}\hat{\Sigma}\hat{\Gamma}^{-1}$. 容易证明 $\hat{\Omega}$ 是 Ω 的相合估计量.

10.4.2 经验似然

如果把 m_1 和 m_2 看成已知, 那么模型 (10.4.2) 可以看成线性模型. 利用线性模型中参数的经验似然比的构造方法可以构造 β_0 的经验似然比函数. 用 m_1 和 m_2 的估计量代替它们即可得到 β_0 的估计的经验似然比函数. 根据这个思想, 可以引入辅助随机变量

$$\hat{\eta}_i(\beta_0) = (\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta_0)\widetilde{X}_i,$$

其中 $\widetilde{X}_i$ 是 $(I-S)\boldsymbol{X}$ 的第 i 行, $\widetilde{Y}_i$ 是 $(I-S)\boldsymbol{Y}$ 的第 i 个元素, S 在 (10.4.4) 式中定义. β_0 的经验对数似然比函数定义为

$$\hat{R}(\beta_0) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\middle| p_i \geqslant 0, \sum_{i=1}^{n}p_i = 1, \sum_{i=1}^{n}p_i\hat{\eta}_i(\beta_0) = 0\right\}.$$

极大化 $\{-\hat{R}(\beta)\}$ 可以得到 β_0 的极大经验似然估计量. 我们有下列渐近结果.

定理 10.4.2 设条件 (C10.4.1)~(C10.4.4) 成立, 则 $\hat{R}(\beta_0) \xrightarrow{D} \chi_p^2$, 其中 χ_p^2 表示自由度为 p 的 χ^2 分布.

使用定理 10.4.2 可以构造 β_0 的近似 $1-\alpha$ 置信域 $\{\beta : \hat{R}(\beta_0) \leqslant \chi_p^2(1-\alpha)\}$, $0<\alpha<1$, 其中 $\chi_p^2(1-\alpha)$ 为 χ_p^2 分布的 $1-\alpha$ 分位数.

10.4.3 变量选择

利用 (10.4.2) 和 (10.4.5) 式, 定义惩罚校正的最小二乘函数为

$$\mathcal{L}_p(\beta) = \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\beta)^{\mathrm{T}}(I-S)^{\star 2}(\boldsymbol{Y} - \boldsymbol{X}\beta) + n\sum_{j=1}^{p}p_{\lambda_j}(|\beta_j|), \tag{10.4.7}$$

其中 $p_{\lambda_j}(\cdot)$ 是惩罚函数, λ_j 是调整参数. 这里取 SCAD 惩罚函数, 其定义见 (1.5.3) 式. 关于 β 极小化 $\mathcal{L}_p(\beta)$ 可得到 β_0 的惩罚校正的最小二乘估计量 $\hat{\beta}_{PC}$.

下面讨论 $\hat{\beta}_{\mathrm{PC}}$ 的渐近性质. 设真值 $\beta_0 = (\beta_{01}, \cdots, \beta_{0p})^{\mathrm{T}} = (\beta_{10}^{\mathrm{T}}, \beta_{20}^{\mathrm{T}})^{\mathrm{T}}$. 不失一般性, 设 β_0 包含所有非零分量, $\beta_{20} = 0$. 用 s 表示 β_{10} 的维数. 记

$$a_n = \max_{1\leqslant j\leqslant p}\{p'_{\lambda_j}(|\beta_{0j}|) : \beta_{0j} \neq 0\},$$

$$b_n = \max_{1\leqslant j\leqslant p}\{p''_{\lambda_j}(|\beta_{0j}|) : \beta_{0j} \neq 0\}.$$

除了条件 (C10.4.1)~(C10.4.4) 外, 我们再给出如下条件

(C10.4.5) 随机变量 (Z_{11}, Z_{12}) 的联合密度函数 $f(z_1, z_2)$ 具有紧支撑且有界, 具有有界连续的二阶偏导数, 且 $\sup\limits_{z_1,z_2}\left|\dfrac{f(z_1,z_2)}{f_1(z_1)f_2(z_2)}\right| < 1$.

我们有下列渐近结果.

定理 10.4.3　设条件 (C10.4.1)~(C10.4.5) 成立. 如果当 $n\to\infty$ 时, $a_n\to 0$, $b_n\to 0$, 则 $\mathcal{L}_p(\beta)$ 的局部极大值 $\hat\beta_{\rm PC}$ 存在, 且

$$||\hat\beta_{\rm PC}-\beta_0||=O_P(n^{-1/2}+a_n).$$

下面的定理给出了惩罚校正的最小二乘估计量的 oracle 性质.

定理 10.4.4(Oracle 性质)　设条件 (C10.4.1)~(C10.4.5) 成立. 又设当 $n\to\infty$ 时, $\lambda_j\to 0$, $\sqrt{n}\lambda_j\to\infty$, $j=1,\cdots,p$, 惩罚函数 $p_{\lambda_j}(|\beta_j|)$ 满足 $\liminf\limits_{n\to\infty}\liminf\limits_{\beta_j\to 0_+}\lambda_j^{-1}p'_{\lambda_j}(|\beta_j|)>0$. 如果 $a_n=O_P(n^{-1/2})$, 则依概率趋于 1, 定理 10.4.3 中 $\sqrt{n}$ 相合的局部极小值 $\hat\beta_{\rm PC}=(\hat\beta_{\rm PC1}^{\rm T},\hat\beta_{\rm PC2}^{\rm T})^{\rm T}$ 满足:

(1) (稀疏性) $\hat\beta_{\rm PC2}=0$;

(2) (渐近正态性)

$$\sqrt{n}[\Gamma_1+\Sigma_\lambda][\hat\beta_{\rm PC1}-\beta_{10}+(\Gamma_1+\Sigma_{\boldsymbol\lambda})^{-1}b]\xrightarrow{D}N(0,\Sigma_1)$$

其中 $\Sigma_{\boldsymbol\lambda}={\rm diag}\{p''_{\lambda_1}(|\beta_{01}|),\cdots,p''_{\lambda_s}(|\beta_{0s}|)\}$, $\boldsymbol{b}=(p'_{\lambda_1}(|\beta_{01}|){\rm sgn}(\beta_{01}),\cdots,p'_{\lambda_s}(|\beta_{0s}|){\rm sgn}(\beta_{0s}))^{\rm T}$, Γ_1 和 Σ_1 分别由 Γ 和 Σ 的 s 行和列组成.

利用 BIC 准则可以选择调整参数, 参看 (Zhou et al., 2011). 具体做法如下. 设

$$l(\beta)=\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\beta)^{\rm T}(I-S)^{\star 2}(\boldsymbol{Y}-\boldsymbol{X}\beta).$$

由 Wang 等 (2007), BIC 得分函数可定义为

$${\rm BIC}(\boldsymbol\lambda)=\log({\rm RSS}_{\boldsymbol\lambda})+e(\boldsymbol\lambda)\log(n)/n,$$

其中 ${\rm RSS}_{\boldsymbol\lambda}=(\boldsymbol{Y}-\boldsymbol{X}\hat\beta)^{\rm T}(I-S)^{\star 2}(\boldsymbol{Y}-\boldsymbol{X}\hat\beta)$, $e(\boldsymbol\lambda)={\rm tr}[(\mathcal{L}''_P(\hat\beta))^{-1}l''(\hat\beta)]$. 选择 $\hat{\boldsymbol\lambda}=\arg\min\limits_{\boldsymbol\lambda}\{{\rm BIC}(\boldsymbol\lambda)\}$.

为了寻找最优的 $\boldsymbol\lambda$, 需要在 d 维空间上极小化 ${\rm BIC}(\boldsymbol\lambda)$. 这是一个相当困难的任务. 因此, 遵照 Fan 和 Li(2004) 的思想, 我们取 $\lambda_j=\lambda{\rm SE}(\hat\beta_j)$, 其中 ${\rm SE}(\hat\beta_j)$ 为非惩罚校正的最小二乘估计 $\hat\beta_j$ 的估计的标准误. 于是, 利用对单参数的极小化方法可以得到调整参数.

10.5　广义可加部分线性模型

考虑一维响应变量 Y 与 $p+d$ 维协变量 (X,Z) 之间的相关性, 设未知的条件均值函数为 $\mu(x,z)=E(Y|X=x,Z=z)$, 条件方差函数为 ${\rm var}(Y|X=x,Z=z)=\sigma^2V(\mu(x,z))$, 其中 $V(\cdot)$ 是已知正函数. 那么, 利用已知的联系函数 $g(\cdot)$ 可将均值函数 $\mu(\cdot)$ 定义为可加部分线性函数:

$$g(\mu(x,z)=\beta_0^{\rm T}x+m_1(z_1)+\cdots+m_d(z_d),\tag{10.5.1}$$

其中 β_0 是 p 维未知参数, 诸 m_j 是未知函数. 为确保诸函数 m_j 的可识别性, 假设 $E[m_j(Z_{1j})]=0,\ j=1,\cdots,d$. 称模型 (10.5.1) 为广义可加部分线性模型.

本节有两个目的: 一是提出一个简单方法来估计模型 (10.5.1) 中的参数和非参数分量; 二是讨论如何在这个半参数框架下选择重要的参数变量. 本节内容主要取材于 Wang 等 (2011).

10.5.1 估计方法

设 $\{(X_i,Z_i,Y_i),1\leqslant i\leqslant n\}$ 是来自 (X,Z,Y) 的 iid 样本, 其中 $X_i=(X_{i1},\cdots,X_{ip})^{\mathrm{T}}$, $Z_i=(Z_{i1},\cdots,Z_{id})^{\mathrm{T}}$, $i=1,\cdots,n$. 不失一般性, 假设 Z_{ij} 在区间 $[0,1]$ 上取值. 在光滑假定下, 诸 $m_j(\cdot)$ 可以用样条函数逼近. 记 $m(z)=m_1(z_1)+\cdots+m_d(z_d)$. 我们将考虑 $m(z)$ 的多项式样条估计 $\hat{m}(z)$. 参看 10.2.3 小节. 仍使用 10.2.3 小节的记号, 用 $\mathcal{G}_j^0$ 表示 $[0,1]$ 上的经验中心化多项式样条空间, $\mathcal{G}$ 表示所有形如 $\widetilde{m}(z)=\widetilde{m}_1(z_1)+\cdots+\widetilde{m}_d(z_d)$ 的可加函数 $\widetilde{m}(z)$ 构成的集合, 其中每一个分量函数 $\widetilde{m}_j\in\mathcal{G}_j^0$. 通过极大化拟似然函数

$$L(\beta,\widetilde{m})=\frac{1}{n}\sum_{i=1}^{n}Q[g^{-1}(\beta^{\mathrm{T}}X_i+\widetilde{m}(Z_i)),Y_i] \tag{10.5.2}$$

来寻找一个 β 的值和一个函数 $\widetilde{m}\in\mathcal{G}$. 对于广义模型, 通常取拟似然函数 $Q(u,y)$ 满足

$$\frac{\partial}{\partial u}Q(u,y)=\frac{y-u}{V(u)}.$$

选定节点数 $N=N_n$, 记 $R=N+p$. 对第 j 个协变量 z_j, 设 $\{B_{rj}(z_j),1\leqslant r\leqslant R\}$ 为 p 阶多项式样条基函数. 对任何 $\widetilde{m}\in\mathcal{G}$, 记 $\widetilde{m}=\gamma^{\mathrm{T}}B(z)$, 其中 $B(z)=\{B_{rj}(z_j),1\leqslant r\leqslant R,1\leqslant j\leqslant d\}^{\mathrm{T}}$ 为样条基函数族, $\gamma=\{\gamma_{rj},1\leqslant r\leqslant R,1\leqslant j\leqslant d\}^{\mathrm{T}}$ 为样条系数向量. 因此, (10.5.2) 的极大化问题等价于对 β 和 γ 极大化

$$l(\beta,\gamma)=\frac{1}{n}\sum_{i=1}^{n}Q[g^{-1}(\beta^{\mathrm{T}}X_i+\gamma^{\mathrm{T}}B(Z_i)),Y_i], \tag{10.5.3}$$

其最大值记作 $\hat{\beta}$ 和 $\hat{\lambda}=\{\hat{\gamma}_{rj},1\leqslant r\leqslant R,1\leqslant j\leqslant d\}^{\mathrm{T}}$. $m(z)$ 的样条估计量为 $\hat{m}(z)=\hat{\lambda}^{\mathrm{T}}(z)B(z)$.

Wang 等 (2011) 在一组条件下证明了 $\hat{\beta}$ 的渐近正态性, 并给出了 $\hat{m}(z)$ 的收敛速度.

10.5.2 变量选择

利用估计量 $\hat{m}(z)=\hat{\lambda}^{\mathrm{T}}(z)B(z)$ 可以构造惩罚拟似然

$$\mathcal{L}_P(\beta)=\sum_{i=1}^{n}Q[g^{-1}(\beta^{\mathrm{T}}X_i+\hat{\gamma}^{\mathrm{T}}B(Z_i)),Y_i]+n\sum_{j=1}^{d}p_{\lambda_j}(|\beta_j|), \tag{10.5.4}$$

其中 $p_{\lambda_j}(|\beta_j|)$ 取为 SCAD 惩罚函数, 其定义见 (1.5.3) 式. 极大化 $\mathcal{L}_P(\beta)$ 可以得到 β_0 的极大惩罚似然估计量, 记作 $\hat{\beta}_{\mathrm{MPL}}$. Wang 等 (2011) 在一些条件下证明了 $\hat{\beta}_{\mathrm{MPL}}$ 的 oracle 性质, 并给出了它的收敛速度.

可以使用 Fan 和 Li(2001) 提出的广义交叉验证 (GCV) 选择未知参数 λ_j. 具体做法如下. 用估计量 $\hat{\beta}_{\mathrm{MPL}}$ 代替 (10.5.2) 式中的 β, 并对 γ 极大化 $l(\gamma, \hat{\beta}_{\mathrm{MPL}})$, 其解记作 $\hat{\gamma}_{\mathrm{MPL}}$, 且 $m(z)$ 的相应估计量定义为 $\hat{m}_{\mathrm{MPL}}(z) = (\hat{\gamma}_{\mathrm{MPL}})^{\mathrm{T}} B(z)$. 因此, GCV 统计量定义为

$$\mathrm{GCV}(\lambda_1, \cdots, \lambda_d) = \frac{\sum_{i=1}^{n} D[Y_i, g^{-1}(X_i^{\mathrm{T}} \hat{\beta}_{\mathrm{MPL}} + \hat{m}_{\mathrm{MPL}}(Z_i))]}{n[1 - e(\lambda_1, \cdots, \lambda_d)/n]^2},$$

其中 $D(y, u)$ 是 Y 相应于拟合 $(\lambda_1, \cdots, \lambda_d)$ 的偏性,

$$e(\lambda_1, \cdots, \lambda_d) = \mathrm{tr}\{[l''(\hat{\beta}_{\mathrm{MPL}}) - n\Sigma_\lambda(\hat{\beta}_{\mathrm{MPL}})]^{-1} l''(\hat{\beta}_{\mathrm{MPL}})\},$$

$$l'(\beta) = \frac{\partial l(\beta, \hat{\gamma})}{\partial \beta}, \quad l''(\beta) = \frac{\partial^2 l(\beta, \hat{\gamma})}{\partial \beta \partial \beta^{\mathrm{T}}},$$

$$\Sigma_\lambda(\beta)\mathrm{diag} = \left\{ \frac{p'_{\lambda_1}(|\beta_1|)}{|\beta_1|}, \cdots, \frac{p'_{\lambda_d}(|\beta_d|)}{|\beta_d|} \right\}.$$

上述关于 d 维空间的极小化问题是困难的. 因此, 遵照 Fan 和 Li(2004) 的方法, 在实际中取 $\lambda_j = \lambda \mathrm{SE}(\hat{\beta}_j)$, 其中 $\mathrm{SE}(\hat{\beta}_j)$ 为 $\hat{\beta}_j$ 的估计的标准误. 于是, 极小化问题可以削减为一维问题, 使用格子点搜索可以估计调整参数.

参考文献

Alimadad A, Salibian-Barrera M. 2011. An outlier-robust fit for generalized additive models with applications to disease outbreak detection. Journal of the American Statistical Association, 106(494): 719～731

Alvarez J D, Pardinas J R. 2009. Additive models in censored regression. Computational Statistics and Data Analysis, 53(9): 3490～3501

Andrews D, Whang Y. 1990. Additive interactive regression models: circumvention of the curse of the dimensionality. Econometric Theory, 6: 466～479

Belitz C, Lang S. 2008. Simultaneous selection of variables and smoothing parameters in structured additive regression models. Computational Statistics and Data Analysis, 53(1): 61～81

Biedermann S, Dette H, Woods D C. 2011. Optimal design for additive partially nonlinear models. Biometrika, 98(2): 449～458

Buja A, Hastiet J, Tibshiranri J. 1989. Linear smoothers and additive models. The Annals of Statistics, 17: 453~555

Cantoni E, Flemming J M, Ronchetti E. 2011. Variable selection in additive models by non-negative garrote. Statistical Modelling, 11(3): 237~252

Debbarh M, Maillot B. 2008. Additive regression model for continuous time processes. Communications in Statistics-Theory and Methods, 37(15): 2416~2432

De Bock K W, Coussement K, Van den Poel D. 2010. Ensemble classification based on generalized additive models. Computational Statistics and Data Analysis, 54(6): 1535~1546

Fahrmeir L, Kneib T. 2009. Propriety of posteriors in structured additive regression models: Theory and empirical evidence. Journal of Statistical Planning and Inference, 139(3): 843~859

Fahrmeir L, Lang S. 2001. Bayesian inference for generalized additive mixed models based on Markov random field priors. Apply Statistics, 50(2): 201~220

Fan J, Hädle W, Mammen E. 1998. Direct estimation of low-dimensional components in additive models. The Annals of Statistics, 26: 943~971

Fan J, Li R Z. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96: 1348~1360

Fan J, Li R Z. 2004. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. Journal of the American Statistical Association, 99: 710~723

Friedman J H, Stuetzle W. 1981. Projection pursuit regression. Journal of the American Statistical Association, 76: 817~823

Hädle W, Sperlich S, Spokoiny V. 2001. Structural tests in additive regression. Journal of the American Statistical Association, 96: 1333~1347

Hastie T J, Tibshiranri J. 1990. Generalized Additive Models. Washington: Chapman and Hall

Horowitz J L, Mammen E. 2004. Nonparametric estimation of an additive model with a link function. The Annals of Statistics, 32(6): 2412~2443

Huang J A, Horowitz J L, Wei F R. 2010. Variable selection in nonparametric additive models. The Annals of Statistics, 38(4): 2282~2313

Huang J Z. 1998. Projection estimation in multiple regression with application to functional ANOVA models. The Annals of Statistics, 26: 242~272

Huang J Z. 2003. Local asymptotics for polynomial spline regression. The Annals of Statistics, 31: 1600~1635

Huang J Z, Yang L. 2004. Identification of nonlinear additive autoregression models. Journal of the Royal Statistical Society, Series B, 66: 463~477

Jiang J C, Li J T. 2008. Two-stage local M-estimation of additive models. Science in

China, Series A, 51(7): 1315～1338

Kim W, Linton O B, Hengartner N. 1999. Acomputationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals. Journal of Computational and Graphical Statistics, 8: 278～297

Liang H, Su H Y, Thurston S W, et al.. 2009. Empirical likelihood based inference for additive partial linear measurement error models. Statistics and Its Interface, 2: 83～90

Liang H, Thurston S W, Ruppert D, et al.. 2008. Additive partial linear models with measurement errors. Biometrika, 95(3): 667～678

Lin L, Cui X, Zhu L X. 2009. An adaptive two-stage estimation method for additive models. Scandinavian Journal of Statistics, 36(2): 248～269

Linton O B. 1997. Efficient estimation of additive nonparametric regression models. Biometrika, 84: 469～673

Linton O B, Hädle W. 1996. Estimating additive regression models with known links. Biometrika, 83: 529～540

Linton O B, Nielsen J P. 1995. Estimating structured nonparametric regression models by the kernel method. Biometrika, 82: 93～101

Liu X, Wang L, Liang H. 2011. Estimation and variable selection for semiparametric additive partial linear models. Statistica Sinica, 21(3): 1225～1248

Mammen E, Linton O, Nielsen J. 1999. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. The Annals of Statistics, 27: 1443～1490

Marra G, Wood S N. 2011. Practical variable selection for generalized additive models. Computational Statistics and Data Analysis, 55(7): 2372～2387

Manzan S, Zerom D. 2005. Kernel estimation of a partially linear additive model. Statistics & Probability Letters, 72(4): 313～322

Müller H G, Yao F. 2008. Functional additive models. Journal of the American Statistical Association, 103(484): 1534～1544

Nielsen J P, Sperlich S. 2005. Smooth backfitting in practice. Journal of the Royal Statistical Society, Series B, 67: 43～61

Opsomer J D. 2000. Asymptotic properties of backfitting estimators. Journal of Multivariate Analysis, 73: 166～179

Opsomer J D, Ruppert D. 1997. Fitting a bivariate additive model by local polynomial regression. The Annals of Statistics, 25(1): 186～211

Opsomer J D, Ruppert D. 1999. A root-n consistent backfitting estimator for semiparametric additive modeling. Journal of Computing and Graphics Statistics, 8: 715～732

Panagiotelis A, Smith M. 2008. Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models. Journal of Econometrics, 143(2): 291～316

Ravikumar P, Lafferty J, Liu H, Wasserman L. 2009. Sparse additive models. Journal of the Royal Statistical Society, Series B, 71(5): 1009～1030

Severini T A, Staniswalis J G. 1994. Quasilikelihood estimation in semiparametric models. Journal of the American Statistical Association, 89: 501～511

Sperlich S, Tjøtheim D, Yang L. 2002. Nonparametric estimation and testing of interaction in additive models. Econometric Theory, 18: 197～151

Stonec C J. 1982. Optimal global rates of convergence for nonparametric regression. The Annals of Statistics, 10: 1040～1053

Stone C J. 1985. Additive regression and other nonparametric models. The Annals of Statistics, 13(2): 689～705

Stone C J. 1994. The use of polynomial splines and their tensor products in multivariate function estimation. The Annals of Statistics, 22: 118～184

Tibshirani R. 1996. Regression shrinkage and selection via the Lasso. Journal of Royal Statistical Society, Series B, 58: 267～288

Tjøtheim D, Auestad B. 1994. Nonparametric identification of nonlinear time series: projections. Journal of the American Statistical Association, 89: 1398～1409

Wang J, Yang L J. 2009. Efficient and fast spline-backfitted kernel smoothing of additive models. Annals of the Institute of Statistical Mathematics, 61(3): 663～690

Wang L, Liu X, Liang H. 2011. Estimation and variable selection for generalized additive partial linear models. The Annals of Statistics, 39(4): 1827～1851

Wang L, Yang L J. 2007. Spline-backfitted kernel smoothing of nonlinear additive autoregression model. The Annals of Statistics, 35(6): 2474～2503

Wang X L, Chen F, Lin L. 2010. Empirical likelihood inference for the parameter in additive partially linear EV models. Communications in Statistics-Theory and Methods, 39: 3513～3524

Wood S N. 2008. Fast stable direct fitting and smoothness selection for generalized additive models. Journal of the Royal Statistical Society, Series B, 70(3): 495～518

Xia Y C. 2009. A note on the backfitting estimation of additive models. Bernoulli, 15(4): 1148～1153

Xue L, Liang H. 2010. Polynomial spline estimation for a generalized additive coefficient model. Scandinavian Journal of Statistics, 37(1): 26～46

Xue L, Yang L. 2006a. Additive coefficient modeling via polynomial spline. Statistica Sinica, 16, 1423～1446

Xue L, Yang L. 2006b. Estimation of semiparametric additive coefficient model. Journal of Statistical Planning and Inference, 136: 2506～2534

Yang L, Hädle W, Nielsen J P. 1999. Nonparametric autoregression with multiplicative volatility and additive mean. Journal of Time Series Analysis, 20: 579～504

Yang L, Park B U, Xue L, et al.. 2006. Estimation and testing of varying coefficients in

additive models with marginal integration. Journal of the American Statistical Association, 101: 1212～1227

Yang L, Sperlich S, Hädle W.2003. Derivative estimation and testing in generalized additive models. Journal of Statistical Planning and Inference, 115: 521～542

You J, Zhou H. 2007. Two-stage efficient estimation of longitudinal nonparametric additive models. Statistics & Probability Letters, 77(17): 1666～1675

Yu K, Park B U, Mammen E. 2008. Smooth backfitting in generalized additive models. The Annals of Statistics, 36(1): 228～260

Yuan M, Lin Y. 2006. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society, Series B, 68: 49～67

Zhang X Y, Liang H. 2011. Focused information criterion and model averaging for generalized additive partial linear models. The Annals of Statistics, 39(1): 174～200

Zheng S R. 2008. Selection of components and degrees of smoothing via lasso in high dimensional nonparametric additive models. Computational Statistics and Data Analysis, 53(1): 164～175

Zhou Z G, Jiang R, Qian W M. 2011. Variable selection for additive partially linear models with measurement error. Metrika, 74(2): 185～202

Zou H. 2006. The adaptive Lasso and its oracle properties. Journal of American Statistical Association, 101: 1418～1429

《现代数学基础丛书》已出版书目

(按出版时间排序)

1 数理逻辑基础(上册) 1981.1 胡世华 陆钟万 著
2 紧黎曼曲面引论 1981.3 伍鸿熙 吕以辇 陈志华 著
3 组合论(上册) 1981.10 柯 召 魏万迪 著
4 数理统计引论 1981.11 陈希孺 著
5 多元统计分析引论 1982.6 张尧庭 方开泰 著
6 概率论基础 1982.8 严士健、王隽骧 刘秀芳 著
7 数理逻辑基础(下册) 1982.8 胡世华 陆钟万 著
8 有限群构造(上册) 1982.11 张远达 著
9 有限群构造(下册) 1982.12 张远达 著
10 环与代数 1983.3 刘绍学 著
11 测度论基础 1983.9 朱成熹 著
12 分析概率论 1984.4 胡迪鹤 著
13 巴拿赫空间引论 1984.8 定光桂 著
14 微分方程定性理论 1985.5 张芷芬 丁同仁 黄文灶 董镇喜 著
15 傅里叶积分算子理论及其应用 1985.9 仇庆久等 编
16 辛几何引论 1986.3 J.柯歇尔 邹异明 著
17 概率论基础和随机过程 1986.6 王寿仁 著
18 算子代数 1986.6 李炳仁 著
19 线性偏微分算子引论(上册) 1986.8 齐民友 著
20 实用微分几何引论 1986.11 苏步青等 著
21 微分动力系统原理 1987.2 张筑生 著
22 线性代数群表示导论(上册) 1987.2 曹锡华等 著
23 模型论基础 1987.8 王世强 著
24 递归论 1987.11 莫绍揆 著
25 有限群导引(上册) 1987.12 徐明曜 著
26 组合论(下册) 1987.12 柯 召 魏万迪 著
27 拟共形映射及其在黎曼曲面论中的应用 1988.1 李 忠 著
28 代数体函数与常微分方程 1988.2 何育赞 著
29 同调代数 1988.2 周伯壎 著

30　近代调和分析方法及其应用　1988.6　韩永生　著
31　带有时滞的动力系统的稳定性　1989.10　秦元勋等　编著
32　代数拓扑与示性类　1989.11　马德森著　吴英青　段海鲍译
33　非线性发展方程　1989.12　李大潜　陈韵梅　著
34　反应扩散方程引论　1990.2　叶其孝等　著
35　仿微分算子引论　1990.2　陈恕行等　编
36　公理集合论导引　1991.1　张锦文　著
37　解析数论基础　1991.2　潘承洞等　著
38　拓扑群引论　1991.3　黎景辉　冯绪宁　著
39　二阶椭圆型方程与椭圆型方程组　1991.4　陈亚浙　吴兰成　著
40　黎曼曲面　1991.4　吕以辇　张学莲　著
41　线性偏微分算子引论(下册)　1992.1　齐民友　著
42　复变函数逼近论　1992.3　沈燮昌　著
43　Banach 代数　1992.11　李炳仁　著
44　随机点过程及其应用　1992.12　邓永录等　著
45　丢番图逼近引论　1993.4　朱尧辰等　著
46　线性微分方程的非线性扰动　1994.2　徐登洲　马如云　著
47　广义哈密顿系统理论及其应用　1994.12　李继彬　赵晓华　刘正荣　著
48　线性整数规划的数学基础　1995.2　马仲蕃　著
49　单复变函数论中的几个论题　1995.8　庄圻泰　著
50　复解析动力系统　1995.10　吕以辇　著
51　组合矩阵论　1996.3　柳柏濂　著
52　Banach 空间中的非线性逼近理论　1997.5　徐士英　李　冲　杨文善　著
53　有限典型群子空间轨道生成的格　1997.6　万哲先　霍元极　著
54　实分析导论　1998.2　丁传松等　著
55　对称性分岔理论基础　1998.3　唐　云　著
56　Gel'fond-Baker 方法在丢番图方程中的应用　1998.10　乐茂华　著
57　半群的 S-系理论　1999.2　刘仲奎　著
58　有限群导引(下册)　1999.5　徐明曜等　著
59　随机模型的密度演化方法　1999.6　史定华　著
60　非线性偏微分复方程　1999.6　闻国椿　著
61　复合算子理论　1999.8　徐宪民　著
62　离散鞅及其应用　1999.9　史及民　编著
63　调和分析及其在偏微分方程中的应用　1999.10　苗长兴　著

64 惯性流形与近似惯性流形 2000.1 戴正德 郭柏灵 著
65 数学规划导论 2000.6 徐增堃 著
66 拓扑空间中的反例 2000.6 汪 林 杨富春 编著
67 拓扑空间论 2000.7 高国士 著
68 非经典数理逻辑与近似推理 2000.9 王国俊 著
69 序半群引论 2001.1 谢祥云 著
70 动力系统的定性与分支理论 2001.2 罗定军 张 祥 董梅芳 编著
71 随机分析学基础(第二版) 2001.3 黄志远 著
72 非线性动力系统分析引论 2001.9 盛昭瀚 马军海 著
73 高斯过程的样本轨道性质 2001.11 林正炎 陆传荣 张立新 著
74 数组合地图论 2001.11 刘彦佩 著
75 光滑映射的奇点理论 2002.1 李养成 著
76 动力系统的周期解与分支理论 2002.4 韩茂安 著
77 神经动力学模型方法和应用 2002.4 阮炯 顾凡及 蔡志杰 编著
78 同调论——代数拓扑之一 2002.7 沈信耀 著
79 金兹堡-朗道方程 2002.8 郭柏灵等 著
80 排队论基础 2002.10 孙荣恒 李建平 著
81 算子代数上线性映射引论 2002.12 侯晋川 崔建莲 著
82 微分方法中的变分方法 2003.2 陆文端 著
83 周期小波及其应用 2003.3 彭思龙 李登峰 谌秋辉 著
84 集值分析 2003.8 李 雷 吴从炘 著
85 数理逻辑引论与归结原理 2003.8 王国俊 著
86 强偏差定理与分析方法 2003.8 刘 文 著
87 椭圆与抛物型方程引论 2003.9 伍卓群 尹景学 王春朋 著
88 有限典型群子空间轨道生成的格(第二版) 2003.10 万哲先 霍元极 著
89 调和分析及其在偏微分方程中的应用(第二版) 2004.3 苗长兴 著
90 稳定性和单纯性理论 2004.6 史念东 著
91 发展方程数值计算方法 2004.6 黄明游 编著
92 传染病动力学的数学建模与研究 2004.8 马知恩 周义仓 王稳地 靳 祯 著
93 模李超代数 2004.9 张永正 刘文德 著
94 巴拿赫空间中算子广义逆理论及其应用 2005.1 王玉文 著
95 巴拿赫空间结构和算子理想 2005.3 钟怀杰 著
96 脉冲微分系统引论 2005.3 傅希林 闫宝强 刘衍胜 著
97 代数学中的 Frobenius 结构 2005.7 汪明义 著

98 生存数据统计分析 2005.12 王启华 著
99 数理逻辑引论与归结原理(第二版) 2006.3 王国俊 著
100 数据包络分析 2006.3 魏权龄 著
101 代数群引论 2006.9 黎景辉 陈志杰 赵春来 著
102 矩阵结合方案 2006.9 王仰贤 霍元极 麻常利 著
103 椭圆曲线公钥密码导引 2006.10 祝跃飞 张亚娟 著
104 椭圆与超椭圆曲线公钥密码的理论与实现 2006.12 王学理 裴定一 著
105 散乱数据拟合的模型方法和理论 2007.1 吴宗敏 著
106 非线性演化方程的稳定性与分歧 2007.4 马 天 汪宁宏 著
107 正规族理论及其应用 2007.4 顾永兴 庞学诚 方明亮 著
108 组合网络理论 2007.5 徐俊明 著
109 矩阵的半张量积:理论与应用 2007.5 程代展 齐洪胜 著
110 鞅与 Banach 空间几何学 2007.5 刘培德 著
111 非线性常微分方程边值问题 2007.6 葛渭高 著
112 戴维-斯特瓦尔松方程 2007.5 戴正德 蒋慕蓉 李栋龙 著
113 广义哈密顿系统理论及其应用 2007.7 李继彬 赵晓华 刘正荣 著
114 Adams 谱序列和球面稳定同伦群 2007.7 林金坤 著
115 矩阵理论及其应用 2007.8 陈公宁 著
116 集值随机过程引论 2007.8 张文修 李寿梅 汪振鹏 高 勇 著
117 偏微分方程的调和分析方法 2008.1 苗长兴 张 波 著
118 拓扑动力系统概论 2008.1 叶向东 黄 文 邵 松 著
119 线性微分方程的非线性扰动(第二版) 2008.3 徐登洲 马如云 著
120 数组合地图论(第二版) 2008.3 刘彦佩 著
121 半群的 S-系理论(第二版) 2008.3 刘仲奎 乔虎生 著
122 巴拿赫空间引论(第二版) 2008.4 定光桂 著
123 拓扑空间论(第二版) 2008.4 高国士 著
124 非经典数理逻辑与近似推理(第二版) 2008.5 王国俊 著
125 非参数蒙特卡罗检验及其应用 2008.8 朱力行 许王莉 著
126 Camassa-Holm 方程 2008.8 郭柏灵 田立新 杨灵娥 殷朝阳 著
127 环与代数(第二版) 2009.1 刘绍学 郭晋云 朱 彬 韩 阳 著
128 泛函微分方程的相空间理论及应用 2009.4 王 克 范 猛 著
129 概率论基础(第二版) 2009.8 严士健 王隽骧 刘秀芳 著
130 自相似集的结构 2010.1 周作领 瞿成勤 朱智伟 著
131 现代统计研究基础 2010.3 王启华 史宁中 耿 直 主编

132 图的可嵌入性理论(第二版) 2010.3 刘彦佩 著
133 非线性波动方程的现代方法(第二版) 2010.4 苗长兴 著
134 算子代数与非交换 L_p 空间引论 2010.5 许全华、吐尔德别克、陈泽乾 著
135 非线性椭圆型方程 2010.7 王明新 著
136 流形拓扑学 2010.8 马 天 著
137 局部域上的调和分析与分形分析及其应用 2011.6 苏维宜 著
138 Zakharov 方程及其孤立波解 2011.6 郭柏灵 甘在会 张景军 著
139 反应扩散方程引论(第二版) 2011.9 叶其孝 李正元 王明新 吴雅萍 著
140 代数模型论引论 2011.10 史念东 著
141 拓扑动力系统——从拓扑方法到遍历理论方法 2011.12 周作领 尹建东 许绍元 著
142 Littlewood-Paley 理论及其在流体动力学方程中的应用 2012.3 苗长兴 吴家宏 章志飞 著
143 有约束条件的统计推断及其应用 2012.3 王金德 著
144 混沌、Mel'nikov 方法及新发展 2012.6 李继彬 陈凤娟 著
145 现代统计模型 2012.6 薛留根 著